

Metabolische Modelle – Aber ohne phylogenetischem Trugschluss!

Die Problematik und Optimierung der Funktionsvorhersage von Enzymen und der Erstellung metabolischer Modelle auf Grundlage der Phylogenie am Beispiel der Rhodobacteraceen und Archaeen.

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Marcus Ulbrich
aus Arnstadt

1. Referent:	Prof. Dr. Dietmar Schomburg
2. Referent:	Prof. Dr. Dieter Jahn
eingereicht am:	25.04.2016
mündliche Prüfung (Disputation) am:	21.06.2016

Druckjahr:	2016
------------	------

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Drüppel, K., Hensler, M., Trautwein, K., Koßmehl, S., Wöhlbrand, L., Schmidt-Hohagen, K., Ulbrich, M., Bergen, N., Meier-Kolthoff, J.P., Göker, M., et al.: **Pathways and substrate-specific regulation of amino acid degradation in *Phaeobacter inhibens* DSM 17395 (archetype of the marine *Roseobacter* clade)**. Environ. Microbiol., 16, 218-238 (2014).

Wiegmann, K., Hensler, M., Wöhlbrand, L., Ulbrich, M., Schomburg, D. and Rabus, R.: **Carbohydrate catabolism in *Phaeobacter inhibens* DSM 17395, member of the marine *Roseobacter* Clade**. Appl. Environ. Microbiol., AEM.00719-14 (2014).

Chang, A., Schomburg, I., Placzek, S., Jeske, L., Ulbrich, M., Xiao, M., Sensen, CW., Schomburg, D.: **BRENDA in 2015: exciting developments in its 25th year of existence**. Nucleic Acids Res., published online doi: 10.1093/nar/gku1068 (2014).

Kleist, S., Ulbrich, M., Bill N., Schmidt-Hohagen K., Geffers R., Schomburg D.: **Dealing with salinity extremes and nitrogen limitation – an unexpected strategy of the marine bacterium *Dinoroseobacter shibae***. Environmental Microbiology, doi:10.1111/1462-2920.13266. (2016).

Posterbeiträge

Ulbrich, M., Schomburg, D.: **EnzymeDetector: Enhancements in 2013 for high throughput applications**. German Conference on Bioinformatics, Göttingen (2013)

*In Erinnerung an meinen Urgroßvater Gerhard Hoffman, der mich wahren Fleiß
in den jüngsten Jahren lehrte und mit Dank an meine liebste Sandra, die diesen
Ehrgeiz heute erträgt.*

Inhaltsverzeichnis

Zusammenfassung	VII
Abstract	IX
1 Einleitung und Hintergrundwissen	1
1.1 Die Systembiologie	1
1.2 Metabolische Netzwerke und Modelle	2
1.3 Die Problematik der Erstellung metabolischer Modelle	3
1.4 Die Problematik des phylogenetischen Ansatzes.....	5
1.5 Die EC-Nummern	6
1.6 Der EnzymeDetector	7
1.7 Die Domäne der Archaea.....	10
1.8 Die Familie der Rhodobacteraceae	13
1.9 Die Sequenz-, Annotations-, Stoffwechselweg- und Metadatenquellen.....	15
1.10 Die „Modelliererwelt“ verbessern? – die Leitthesen.....	19
1.11 Die Zielsetzungen der Arbeit.....	21
2 Material und Methoden	22
2.1 Die EC-Nummern	22
2.2 Weitere bioinformatische Programme	22
2.3 Die phylogenetischen Applikationen	23
2.4 Die Sequenz-, Annotations-, Stoffwechselweg- und Metadaten	24
2.5 Das High Performance Computing (HPC)-Cluster	30
2.6 Der EnzymeDetector	31
2.7 Das EnzymeDetector-Webinterface.....	33
2.8 Weitere Software.....	33
2.9 Die Auswerteverfahren metadatentypischer Enzyme/Stoffwechselwege	34
3 Ergebnisse und Diskussion	40
3.1 Die Verbesserungen und Erweiterungen des EnzymeDetector.....	40
3.2 Das „Pathwaylayout“ mit MetaCyc	48
3.3 Die Analysegruppen.....	50
3.4 Die Enzym-Evolution der Archaeen.....	50
3.5 Die Enzym-Matrizen (Metadaten)	75
3.6 Die Stoffwechselweg-Matrizen (Metadaten).....	83
3.7 Die gruppentypischen Lücken in Stoffwechselwegen	90
3.8 Neue gruppentypische Stoffwechselwegansätze	93

3.9 Das Core-/Pan-„Enzynome“	97
3.10 Die phylogenetischen Bäume auf Basis binärer Enzymvektoren.....	101
3.11 Die phylogenetischen Bäume auf Basis binärer Pathway-Vektoren	109
4 Fazit und Ausblick	115
Danksagung	121
Symbole und Abkürzungsverzeichnis	122
Abbildungsverzeichnis	123
Tabellenverzeichnis	124
Anhang	125
<i>Anhang 1: Datentabellen des EnzymeDetector-Ansatzes der Rhodobacteraceen.</i>	<i>125</i>
<i>Anhang 2: Datentabellen des EnzymeDetector-Ansatzes der Archaeen.....</i>	<i>125</i>
<i>Anhang 3: Datentabellen des BLAST-Ansatzes der Archaeen.</i>	<i>125</i>
<i>Anhang 4: Datentabellen des BLAST-Ansatzes der Archaeen zur Auswertung der E-Value-Statistik.</i>	<i>125</i>
<i>Anhang 5: Quellcode des EnzymeDetector, des Pathway-Layout, der Skripte und die Datenbanken</i>	<i>125</i>
Literaturverzeichnis	126
Lebenslauf	155

Zusammenfassung

Die Erstellung metabolischer Modelle infolge der Rekonstruktion von Netzwerken enzymkatalysierter Reaktionen des Stoffwechsels eines Organismus ist ein nicht mehr wegzudenkendes systembiologisches Werkzeug. Die Systembiologie verfolgt den Ansatz, mit Hilfe dieser Modelle und Methoden Organismen und ihre metabolischen Anpassungen, sowohl an äußere Begebenheiten, als auch an organismusspezifische Abläufe, in ihrer Gesamtheit zu simulieren, nachzuvollziehen und zu analysieren.

Eine entscheidende Ausgangsbasis schafft die Enzymfunktionsvorhersage durch die Annotation von Enzymklassen zu den entsprechenden Gensequenzen, denen wiederum Reaktionen zugeordnet werden und so metabolische Abläufe rekonstruierbar sind. Hierfür stehen vielfältige Datenbanken und Methoden der Bioinformatik zur Verfügung, die offensichtlich nur in ihrer Gesamtheit ein der Wirklichkeit entsprechendes Abbild schaffen können. Um möglichst unkompliziert und schnell fundamentierte Aussagen über den Enzympool eines Organismus treffen und diese Informationen im Sinne der metabolischen Modellerstellung verwenden zu können, wird in der vorliegenden Arbeit das Programm EnzymeDetector weiterentwickelt. Somit kann durch die Aggregation der Informationen von automatisierten Annotationsdatenbanken wie dem NCBI, KEGG, PATRIC und manuell geprüfter, bzw. experimentell beschriebener Enzymfunktionen der BRENDA und SwissProt-Datenbank eine umfassende Sammlung genomweiter Funktionsvorhersagen von über 3000 sequenzierten Organismen erstellt werden. Unterstützung bei der Entscheidung für die wahrscheinlichste Genfunktion bieten weitere entwickelte Algorithmen auf Basis von Sequenzvergleichen mittels BLAST, Pfam-Proteinfamilien und Sequenzmustern der BrEPS-Datenbank.

Die Erstellung von Modellen und die Darstellung des Metabolismus ganzer phylogenetischer Kladen und Domänen werden durch dieses Programm erleichtert und es werden in dieser Arbeit am Beispiel der Rhodobacteraceen (Bakterien) und der Archaeen Zusammenhänge der Enzymausstattung zu diversen mikrobiellen Eigenschaften und Metadaten wie dem Habitat der Organismen und dessen Umweltbedingungen interpretiert. Zusätzlich werden diese Aussagen im Kontext ganzer Stoffwechselwege der MetaCyc-Datenbank analysiert. Hierzu wird ein weiteres und neues Programm entwickelt, welches definitive Aussagen durch die Bestimmung von enzymatischen Schlüsselfunktionen und Alternativreaktionen von 1560 Stoffwechselwegen zulässt. Die Interpretationen der Ergebnisse der Verfahren dieser Arbeit analysieren einerseits durch den Vergleich der phylogenetischen Verwandtschaftsverhältnisse auf Grundlage der 16S rRNA und der Distanzen der jeweiligen Enzym- und Stoffwechselwegpools kritisch die Diversität des Metabolismus der untersuchten Organismengruppen und somit die Basis einer weit verbreiteten Methode der Funktionsvorhersage ohne expliziten Sequenznachweis. Andererseits werden domänentypische und mit mikrobiellen Eigenschaften verknüpfte Enzymklassen beschrieben, die regulatorische, sensorische, oder strukturelle Funktionen und Mechanismen katalysieren, bzw. im Kontext metabolischer Zusammenhänge den Bezug der Organismen zu ihrem jeweiligen Habitat und ihren Metadaten charakterisieren und differenzieren.

Abstract

Generating metabolic models in the wake of reconstructing metabolic networks of enzyme catalyzed metabolic reactions of an organism is a highly important tool in Systems Biology. With these models the Systems Biology follows the approach to simulate, reconstruct and analyze organisms and their metabolic adaptations to the environment as well as the organism-specific processes.

The enzyme function prediction offers an important basis by annotating enzyme classes to the respective sequences. Metabolic reactions are allocated to these sequences which enables the reconstruction of metabolic processes. Multiple databases and bioinformatic methods are available for the reconstruction but can only offer an image of the reality in their entirety. In order to predict the enzyme pool of an organism as fast and precise as possible and in order to use these informations in the sense of metabolic model reconstruction the tool EnzymeDetector is further developed in the present thesis. An extensive collection of genome-wide function predictions of more than 3000 sequenced organisms can be created by aggregating the information of automated annotation databases like NCBI, KEGG, PaTRIC and manually revised or experimentally described functions from BRENDA or SwissProt. Further developed algorithms based on sequence comparisons with BLAST, Pfam and BrEPS support the decision for the most probable enzyme function.

The EnzymeDetector simplifies the building of models and the representation of the metabolism of phylogenetic clades and domains. Furthermore, the connection between an organism's enzyme pool and the diverse microbial characteristics and metadata like the habitat and its environmental properties are interpreted. Additionally, these information is analyzed in the context of whole pathways of the MetaCyc database. For these analyses a new program is developed which permits conclusions by determining enzymatic key functions and alternative reactions of 1560 pathways. The interpretations of the results gained by the methods of this thesis analyze the diversity of the organisms' metabolism and the base for the wide spread method of function prediction without explicit sequence evidence by comparison of the phylogenetic relationship based on the 16S rRNA and the distances of the enzyme and pathway pools. Moreover, enzymes are described which are either domain specific or connected to microbial properties and which catalyze regulatory, sensoric, or structural functions and mechanisms and characterize or differentiate the relationship of the organisms and their respective habitat and metadata in a metabolic context.

1 Einleitung und Hintergrundwissen

1.1 Die Systembiologie

Die Biowissenschaften und vor allem die Biotechnologie lösten im Laufe der Jahre einige Probleme der Menschheit. So konnten Arzneimittel gegen teils tödliche Krankheiten ausgelöst durch Pathogene gefunden oder Biostoffe in kolossalen Mengen zum Beispiel für die Ernährung und Nahrungsergänzung produziert werden. Weiterhin sind die Grundsteine und Beiträge dieser Wissenschaft zur Lösung der umweltfreundlichen Energieproblematik der Zukunft aktueller denn je. Ein schon gegenwärtiges Problem der Menschheit stellt indes das Thema der Müllverbreitung von PET (Polyethylenterephthalat) dar. Waren bisher nur einige wenige Pilze bekannt, die sehr langsam diesen Kunststoff abzubauen im Stande sind, ist erst kürzlich ein Bakterium namens *Ideonella sakaiensis* 201-F6 entdeckt worden. [1] Mit Hilfe zweier Enzyme ist dieser Organismus in der Lage, PET zu zersetzen. Das Verständnis der biochemischen Zusammenhänge dieses Lebewesens ist Voraussetzung für eine biotechnologische Optimierung des Organismus, so dass das Bakterium bei der Sanierung PET-belasteter Flächen und Gewässer in der Zukunft ebenso hilfreich sein kann wie beim Recycling Erdöl-basierter Rohstoffe. Dieser Sachverhalt und die eingangs genannten Aspekte verdeutlichen die Wichtigkeit und den Wert der Erkenntnisse auf dem Gebiet der Grundlagenforschung der Biologie und weiterführender Untersuchungen mikrobieller Organismen.

Die Systembiologie ist ein Teilgebiet der Biowissenschaften, welches sich mit dem Metabolismus von Organismen beschäftigt. Dieser stellt in der Gesamtheit aller ablaufenden biochemischen Prozesse direkt den Phänotyp der betrachteten Lebewesen dar, während er physiologisch relevante Abhängigkeiten von mikrobiellen Eigenschaften und Metadaten wie zum Beispiel der Temperatur berücksichtigt. [2] Infolgedessen ist mittlerweile dieses Verständnis unter anderem von enormer Bedeutung in der Medizin zur Identifikation von prognostischen und diagnostischen Indikatoren, den sogenannten Biomarkern [3], und wegweisend für den gezielten Entwurf von Wirkstoffen, um Pathogene und ihren Stoffwechsel zu verstehen und letztendlich Schlüsselfunktionen zu blockieren (*drug design*) [4]. Weiterhin liefern die Methoden und Erkenntnisse der Systembiologie Hinweise für gezielte genetische Eingriffe, um den Stoffwechsel von Organismen dahingehend zu manipulieren, dass technisch nutzbare Produkte, bzw. gegenüber den natürlich produzierten Stoffen veränderte Substanzen mit neuen Eigenschaften und in optimierten Mengen für den kommerziellen Einsatz entstehen (*metabolic engineering*) [5]. In jedem Fall werden fundamentale Kenntnisse für die Grundlagenforschung der Biologie und Daten für die Untersuchung der Phylogenie und Evolutionsgeschichte der Lebewesen gewonnen. [6,7]

Der Metabolismus von Organismen wird zum besseren Verständnis in einzelne Stoffwechselwege unterteilt. Diese bestehen aus einer linearen, verzweigten oder zyklischen Sequenz von biochemischen Reaktionen, die zumeist durch enzymatische Funktionen katalysiert und unter den gegebenen Bedingungen ermöglicht werden. An diesen Stoffumwandlungen ist immer eine definierte

Menge an Metaboliten beteiligt, so dass ebenso bestimmte Stoffmengen letzten Endes produziert werden. Die Metabolite bezeichnen alle chemischen Verbindungen unabhängig von der Reaktionsrichtung der biochemischen Umwandlungen (Substrat/Edukt - Produkt), in denen sie involviert sind. [2] Neben den Stoffumwandlungen existieren auch Reaktionen und Mechanismen, die der Strukturgebung komplexer Stoffgefüge oder dem Stofftransport und der Signalweiterleitung innerhalb des Systems dienen.

Die exakte Definition der enthaltenen Reaktionen von den Substraten bis hin zu den entsprechenden Endprodukten eines Stoffwechselweges ist nicht immer eindeutig. Wie zuvor erwähnt, sind derartige Unterteilungen des Metabolismus frei durch die Wissenschaft vorgenommen, wobei der Stoffwechsel immer als Gesamtsystem ohne feste zeitliche und örtliche Grenzen gesehen werden muss. Allerdings wird nur so die Möglichkeit geschaffen, einzelne Prozesse zu verstehen und modelartig zusammenzusetzen. Von daher können die Länge und die konkreten Einzelreaktionen von Stoffwechselwegen unterstützt durch Enzyme sich von Literaturquelle zu Literaturquelle unterscheiden. Dabei beschreibt die Länge die definierte Anzahl an biochemischen Reaktionsschritten, um die Ausgangsstoffe über entsprechende Zwischenprodukte zu den Endprodukten umzuwandeln. Daneben existieren ebenso alternative Reaktionsketten, die in konformen Produkten resultieren, wie verschiedene Enzymklassen, die diese katalysieren.

Die Gesamtheit aller biochemischen Reaktionen und infolgedessen die Umwandlung von Metaboliten in einem fest definierten Verhältnis ergeben ein metabolisches Netzwerk, in dem die Produkte der einen Stoffwechselwege die Ausgangsstoffe für wiederum andere liefern und somit die Abläufe unter den gegebenen Bedingungen nachzuvollziehen sind. [8] Das so geschaffene Modell entspricht einem Abbild der Wirklichkeit des Metabolismus eines Organismus unter festen Rahmenbedingungen und dient der möglichst genauen Darstellung der physiologischen und biochemischen Eigenschaften und Abläufe der Zellen. [9,10]

1.2 Metabolische Netzwerke und Modelle

Ein metabolisches Modell entspricht vereinfacht dargestellt einem Informationsspeicher, der möglichst alle bekannten Daten zum Stoffwechsel eines zu untersuchenden Organismus aus diversen Quellen wie beispielsweise bioinformatischen Datenbanken, Primärliteratur oder experimentellen Ergebnissen zusammenträgt. [11]

Durch definierte Annahmen können mit Hilfe eines stöchiometrischen Modells quantitative Vorhersagen der Stoffumsetzungen, das heißt der Flüsse durch dieses metabolische Netzwerk aufgestellt werden. [12] Zu den angesprochenen Annahmen zählt einerseits die Zustandsbetrachtung des sogenannten Fließgleichgewichts (*steady state*). Das bedeutet, dass die Menge der jeweiligen Metabolite im System über die Zeit konstant bleibt und der Zu-, bzw. Abtransport von Stoffen und Verbindungen über die Systemgrenzen hinaus einen Nettounterschied nahe 0 aufweist. [2,13] Des Weiteren wird angenommen, dass die Evolution der Organismen den Stoffwechsel dahingehend optimiert, in Abhängigkeit der vorhandenen Nährstoffe, die maximal mögliche Menge an

Biomasse zum Beispiel in Form von höheren Kohlenstoffverbindungen zu produzieren [14].

Essentielle Reaktionen und entsprechende enzymatische Schlüsselfunktionen dieser Schritte können folglich durch die Analyse der Flüsse und deren Verteilung im Netzwerk in konkreten Szenarien, zum Beispiel bei nur einer verfügbaren Kohlenstoffquelle, bestimmt werden. Diese Methode dient dem Verständnis des untersuchten Metabolismus ebenso wie weitere Anwendungen zur Betrachtung der Zusammenhänge des Netzwerks. So kann *in-silico* nachvollzogen werden, welche Reaktionsflüsse durch das Verhindern der Expression bestimmter Gene (*gene-knockout*) blockiert werden und welche alternativen Reaktionswege der Metabolismus vorsieht oder ob der Organismus seine Überlebensfähigkeit verliert. Dies dient zum Beispiel bei pathogenen Mikroorganismen der Vorhersage potentieller Angriffspunkte für Arzneimittel. [15] Ferner kann die Biotechnologie diese Zusammenhänge interpretieren, um die Produktausbeute entsprechender Mikroorganismen in ihrem Sinne zu optimieren. [16]

1.3 Die Problematik der Erstellung metabolischer Modelle

Zu Beginn der Erstellung eines metabolischen Modells zu einem ausgewählten Organismus stehen dem Modellierer verschiedene Arbeitsweisen zu Verfügung. Grundsätzlich wird zwischen dem sogenannten *top-down*-Ansatz und der *bottom-up*-Methode unterschieden. [17–19] Dabei wird auf der einen Seite von einem möglichst automatisiert zusammengetragenen und entsprechend umfangreichen Rohmodell ausgegangen, welches anschließend weiter analysiert und überprüft werden muss. Währenddessen geht man auf der anderen Seite beim Verfolgen des *bottom-up*-Ansatzes von einem kleineren Kernmodell zumeist des Zentralstoffwechsels aus und erweitert dieses anschließend schrittweise um weitere Stoffwechselwege mit Belegen aus der Literatur und experimentellen Erkenntnissen. Vorteilhaft ist hierbei das Betrachten aller hinzugefügten Reaktionen im Kontext des Gesamtmetabolismus. So werden Lücken durch fehlende Metabolite oder enzymatische Reaktionen in Stoffwechselwegen ohne Beleg auf dem Genom des untersuchten Organismus schnell erkannt und können vorläufig durch postulierte Reaktionen aus den entsprechenden bioinformatischen Stoffwechselwegdatenbanken wie BRENDA [20], MetaCyc [21] oder auch KEGG [22] (Kapitel 1.9) überbrückt werden. Quantitativ fehlende oder überschüssige Metabolite können durch vorausgesetzte Transporter oder Austauschreaktionen des Systems ausgeglichen werden. Allerdings nachteilig ist hierbei der nicht zu unterschätzende Zeitaufwand, um etwaige Stoffwechselwege mit dem Organismus zu verknüpfen und vor allem individuelle Abläufe des untersuchten Metabolismus zu erkennen, insofern diese überhaupt schon in der Literatur beschrieben sind.

Hingegen werden beim gegenübergestellten Verfahren *top-down* alle dem Organismus durch diverse Methoden zugewiesene Enzymklassen den katalysierten Reaktionen zugeordnet, um ein umfangreiches Sammelsurium an möglichen Abläufen des Metabolismus zu erhalten. Die entsprechenden Reaktionen werden folglich ebenfalls aus den genannten Datenbanken extrahiert und müssen nun im Nachgang im metabolischen Kontext betrachtet und überprüft werden. Als Grundlage dient allerdings die Genomannotation, das heißt

die Zuordnung von Enzymfunktionen zu den einzelnen Genen, die diese auf dem Genom des Organismus codieren. [14,23] Auf diesen Enzymen aufgebaut resultieren gleichstehende katalysierte Reaktionen, die wiederum das Netzwerk, bzw. das metabolische Modell aufbauen. Dieses so erhaltene Grundmodell deckt zum größten Teil die Stoffwechselwege des modellierten Organismus ab. Die Qualität dieser Abdeckung hängt allerdings direkt von der Güte der Genannotation und der organismusspezifischen Informationen der Quelldatenbanken ab. Neben dem bis hier hin schnellen und zumeist automatisierten Verfahren kann auf eine manuelle Überprüfung zur Identifizierung von den bereits erwähnten Stoffwechselweglücken nicht verzichtet werden. Das größere Problem stellt jedoch das Erkennen und Verifizieren von falschen oder unzureichenden Annotationen von enzymatischen Funktionen dar, bauen doch auf ihnen das resultierende Netzwerk und die Zusammenhänge des Modells auf. Eine entsprechend fundierte und verlässliche Genannotation ist folgerichtig unumgänglich, um eine optimale Ausgangslage für die Modellierung eines Organismus zu erhalten.

Die letztendliche Genauigkeit des erstellten Modells ist, wie erwähnt, bedingt durch die Vollständigkeit und Qualität der Annotation der Enzymfunktionen zu den einzelnen Genabschnitten, genannt Genloci. Je vollkommener diese ist, desto größer ist die Erwartung, den Metabolismus in seiner Gesamtheit darstellen zu können. Umso mehr von korrekten und verlässlichen Zuweisungen ausgegangen werden kann, umso höher ist die Wahrscheinlichkeit, dass reale Abläufe, die tatsächlich den Stoffwechsel des Organismus beschreiben, in das erstellte Modell einfließen und in Zusammenhang gestellt werden können.

Eine Möglichkeit, Gensequenzen zu annotieren, ist es, übereinstimmende Homologien zu experimentell bestätigten und schon bekannten Funktionen ausfindig zu machen. Hierfür stellt die Bioinformatik diverse Datenbanken zur Verfügung, die dafür Vorhersagen treffen (Kapitel 1.9). Dabei ist allerdings zu beachten, dass je mehr homologe Sequenzen existieren und in den Datenbanken bekannt sind, eine umso verlässlichere Funktionsvorhersage getroffen werden kann. Weniger ausführlich untersuchte Organismen und eventuelle individuelle Funktionen können somit aus vereinzelt Quellen nur schwer oder ohne große Verlässlichkeit beschrieben werden. Infolgedessen kann die Kombination mehrerer Vorhersagemethoden derartige Genfunktionen überhaupt erst erkennen und somit Teile des Stoffwechsels offenbaren.

Falsche Funktionsannotationen wirken sich wie beschrieben direkt auf die Genauigkeit des vorhergesagten Phänotyps des Modells negativ aus. Diesbezüglich ist also die Konfidenz der Enzymvorhersage ebenso grundlegende Voraussetzung wie die Vollständigkeit. Dies wirft die Problematik mehrfacher enzymatischer Funktionen eines Gens auf. Oft wird in den Datenbanken nur eine für offensichtlich eingestufte Genfunktion aufgeführt und resultiert demnach in einem Informationsverlust für die Erstellung eines Modells. Tatsächlich aber können einzelne Sequenzabschnitte mehrere Aufgaben übernehmen und für verschiedene Enzymkatalysatoren codieren. Dies ist der Fall bei bi-, bzw. multifunktionellen Enzymen, die faktisch nacheinander codiert voneinander verschiedene Reaktionen katalysieren (Kapitel 1.6), oder bei Enzymklassen, die nicht hochspezifisch sind, da sie zum Beispiel diverse Cofaktoren und Substrate nutzen (Kapitel 1.5).

Dem Modellierer stehen verschiedene bioinformatische Datenbanken für Genannotationen zur Verfügung. Allerdings treffen diese nur in relativ wenigen Fällen konforme Aussagen zu einzelnen Genen. Zum Beispiel sagen die Datenbanken KEGG [24], NCBI [25,26] und PEDANT [27] nur in etwa einem Drittel übereinstimmende Enzymklassen für die gleichen Sequenzen voraus. [23] Eine manuelle Integration der verfügbaren Datenbanken stellt allein durch die Problematik der Abbildung der vorliegenden Annotationen auf übergreifende Genbezeichnungen einen immensen zeitlichen Aufwand dar. Ebenso kann folglich wie erwähnt eine für die Modellierung von Organismen optimale Annotation nur durch die Aggregation diverser Datenbanken und Vorhersagemethoden erhalten werden. Wesentliche Unterstützung in diesem Sinne bietet das in dieser Arbeit erweiterte Programm namens EnzymeDetector, welches Annotationsdatenbanken integriert, Enzymvorhersagen durch eigene Methoden erweitert und die wahrscheinlichsten Funktionen eines Gens mit entsprechender Konfidenz bewertet (Kapitel 1.6).

1.4 Die Problematik des phylogenetischen Ansatzes

Das vorherige Kapitel macht auf die Problematik der Modelliererwelt, die wahrheitsgemäßen Funktionen der einzelnen Gene des Organismus zu kennen, um sie in Beziehung in einem Modell zu setzen, aufmerksam. Dabei fällt es vor allem schwer, bei bisher weniger experimentell untersuchten Organismen verlässliche Aussagen zu treffen. Sei es auf Grund von finanziellen und zeittechnischen Faktoren, Problemen bei der experimentellen Kultivierung oder letztendlich falschen Annahmen, versucht man, mit diversen Methoden diese Informationslücken zu schließen. Werden mittels bioinformatischer Analysen und Vorhersagen allenfalls vage Hinweise für einen konkreten Sequenznachweis gefunden, resultiert dies oftmals in Querverbindungen zu Funktionen aus bereits erstellten und etablierten Modellen anderer mehr oder weniger nah verwandter Organismen. Ein oft verwendetes Beispiel für angeführte Vergleiche und Rückschlüsse sind Modelle von Stämmen des intensiv untersuchten, fakultativ anaeroben Darmbakteriums *Escherichia coli* [28].

Allerdings bleibt bei diesem Verfahren unberücksichtigt, dass nachgewiesen ist, dass ökologische Umweltaspekte, wie etwa das Vorhandensein von Sauerstoff, fundamentalen Einfluss auf den Metabolismus von Organismen haben. [2,14,29] Die Verfügbarkeit beispielsweise von verschiedenen Kohlenstoffquellen, Stickstoffquellen und Sauerstoff differenziert das Reaktionsnetzwerk, welches Organismen oder Organismengruppen im Laufe der Evolution etabliert haben. Infolgedessen ist die Frage zu stellen, ob derartige Voraussetzungen dafür geeignet sind, Vergleiche des Metabolismus anzustellen und ohne die Übereinstimmung oder auch die Kenntnis der Rahmenbedingungen Feststellungen zu übertragen. Oder spielen das Habitat der Organismen mit all seinen Umweltbedingungen und die mikrobiellen Eigenschaften, zusammengefasst als sogenannte Metadaten, eine tragendere Rolle, als diese Methode zulässt?

Weiterhin besteht möglicherweise auch bei eng verwandten phylogenetischen Kladen, das heißt bei monophyletischen Gruppen einer biologischen systematischen Einheit, die auf einen gemeinsamen Vorfahren in der Evolution zurückzuführen ist [14], die Tatsache eine Rolle, dass verschiedene Strategien

entstanden sein können, um auf gleichartige Umwelteinflüsse zu reagieren und sich anzupassen. Führen derartige vergleichende Methoden folglich zu einem fundamentalen Trugschluss der Modellierung von Organismen und können auf Basis der Metadaten eventuell fundiertere Äquivalenzen aufgezeigt werden?

1.5 Die EC-Nummern

Die Zuordnung von Enzymfunktionen und katalysierten Reaktionen zu den sogenannten „Enzyme Commission (EC)“-Nummern ermöglicht eine numerische Klassifizierung und Kategorisierung der Enzyme. Eine Kontrolle der Benennungen und Zuordnungen obliegt einer internationalen Organisation auf dem Gebiet der Biochemie und Molekularbiologie namens *International Union of Biochemistry and Molecular Biology (IUBMB)* [30]. Im Sinne der Kategorisierung handelt es sich um 4-stellige, durch einen Punkt voneinander getrennte Nummern. Die erste Position unterscheidet in die Hauptklassen der Oxidoreduktasen, Transferasen, Hydrolasen, Lyasen, Isomerasen und Ligasen und lässt eine Differenzierung nach der katalysierten Reaktionsart zu. Die folgenden Nummernpositionen kategorisieren weitere Informationen zum beschriebenen Enzym. So wird nach dem Angriffspunkt des Enzyms (Bindungsart), den Nebenprodukten der Reaktion und nach den erforderlichen Reaktionspartnern oder Cofaktoren unterschieden. Cofaktoren stellen diverse Moleküle und Molekülgruppen, die für die aktive Funktion der Enzyme unerlässlich sind, dar. [2]

Nicht zwangsweise müssen alle 4 Positionen zum Zeitpunkt der Erstellung einer EC-Nummer eindeutig zugewiesen sein. Dies ist zum Beispiel bei noch nicht vollständig kategorisierten Reaktionen, bei denen jedoch die Enzymaktivität nachgewiesen ist, der Fall. Ist folglich die katalytische Aktivität des Proteins noch nicht exakt beschrieben, wird dies durch einen Bindestrich kenntlich gemacht. [30] Zum Beispiel steht die EC 3.4.24.- für die Hydrolyse (Hauptklasse „3“) einer Peptidbindung (Unterklasse „4“) während die enzymatische Aktivität von einem Metallion wie dem Zink-Ion Zn^{2+} abhängig (Untergruppe „24“, Metalloendopeptidase), jedoch die genaue enzymatische Funktion und Reaktion noch nachzuweisen und einzuordnen sind. [31] Einige Datenbanken führen indessen für bereits klassifizierte und eingehend beschriebene Enzymklassen, die allerdings noch nicht in der IUBMB Liste geführt, aber vorgeschlagen sind, eigens vergebene Laufnummern. Dementsprechend sind beispielsweise in der BRENDA-Datenbank sogenannte B-Nummern zu finden [20] und in der UniProt werden Enzyme in diesem Fall mit einem „n“ und der Laufnummer kenntlich gemacht [32] (Beschreibung der Datenbanken: Kapitel 1.9).

1.6 Der EnzymeDetector

Die Problematik der Erstellung metabolischer Modelle auf Grundlage von Enzymannotationen aus Kapitel 1.3 weist darauf hin, dass verschiedene Methoden der bioinformatischen Funktionsvorhersage durchaus Differenzen aufweisen können. Dem Modellierer obliegt es nun, die enzymatische Funktion des einzelnen Genlocus zu definieren, die am wahrscheinlichsten ist und somit einen fundamentierten Baustein des zu erstellenden Netzwerkes an Reaktionen darstellt. Dabei strebt der wissenschaftliche Alltag auf diesem Gebiet nach einer möglichst schnellen, eindeutigen und unkomplizierten, jedoch auch nachvollziehbaren Methodik. Der in dieser Arbeit erweiterte EnzymeDetector ist ein Programm, welches bei dieser automatisierten Enzymvorhersage [23] unterstützt und die Grundlage für derartige Zusammenstellungen von Enzymklassen, die auf dem Genom eines Organismus codiert sind, schafft. Dieser Algorithmus wurde in der Arbeitsgruppe des Instituts für Bioinformatik und Biochemie der TU Braunschweig entwickelt und von Zeit zu Zeit weiter elaboriert. Die Kombination der Genannotationen verschiedener Quellen diverser Methoden zur Funktionsvorhersage ist wesentlicher Grundgedanke des Programms. Insofern resultiert der gewichtete Vergleich der Annotationsquellen und der jeweiligen Vorhersage in einer Entscheidungshilfe für die Funktionszuweisung zu einzelnen Genloci. Ein Genlocus beschreibt einen physischen, positionsmäßig festgelegten Ort auf dem Genom, welcher durch die differenzierte Translation zum entsprechend codierten Produkt, bzw. zur Funktion des Gens führt. [33] Das Kriterium für diese Entscheidung zur wahrscheinlichsten Funktion eines Gens liefert ein entsprechender Konfidenzwert der zugewiesenen Enzymklasse. Dieser setzt sich zusammen aus den Einzelbewertungen der aggregierten Datenbanken und kann mit den Standardeinstellungen zwischen 1 und 152 im Falle einer vollständig kongruenten Vorhersage aller Datenbanken und Methoden liegen. Die Standardbewertungen der einzelnen Quellen wurden mit Hilfe einer statistischen Analyse der Genauigkeit der jeweiligen Genannotationen von 12 Referenzorganismen im Vergleich zu einem sogenannten festgelegten „Standard of Truth“ bestimmt. In diesem Zusammenhang dienen die Enzymzuweisungen der SwissProt-Datenbank (Kapitel 1.9) als Referenz für die Richtigkeit bei dem Vergleich und der Auswertung der Annotationsdaten. [23]

Entsprechend abgestufte Konfidenzwerte zwischen dem Minimal- und Maximalwert einer Enzymannotation spiegeln ein Maß für die Wahrscheinlichkeit, dass es sich tatsächlich um die korrekte Vorhersage handelt, wieder. Umso mehr verschiedene Quellen und vor allem Methoden in ein und derselben enzymatischen Funktion einer Gensequenz resultieren, umso höher ist demzufolge die Richtigkeit dieser Zuweisung einzustufen. Dementsprechend kann für jedes Gen eines Organismengenoms eine fundamentierte Vorhersage automatisiert ausgewertet und getroffen werden, um ein metabolisches Netzwerk zu erstellen. Im Sinne der Vollständigkeit und der Korrektheit des immens wichtigen Ausgangspunktes für derartige Modelle und Netzwerke von metabolischen Reaktionen auf Grundlage annotierter Enzyme des betrachteten Organismus, ist ein Vergleich mehrerer Datenbanken und Quellen unumgänglich. Dies unterstreicht eine Statistik, die verdeutlicht, dass verschiedene Annotationsdatenbanken unterschiedlicher Methodik nur in etwas mehr als 20% der Vorhersagen übereinstimmen. [23] Diese Unterschiede heben

die Bedeutung eines Programms wie dem EnzymeDetector hervor. Ohne eine Einstufung der Konfidenz von Enzymzuweisungen fällt jedem Anwender die qualitative Unterscheidung der Annotationen schwer. Eine Begründung für definitive Entscheidungen der Genfunktionen scheint schier unmöglich, während jede Quelle ebenso einen individuellen Beitrag zur Gesamtvorhersage des Genoms beisteuern kann und infolgedessen eine zufriedenstellende Abdeckung aller Informationen möglich ist.

Neben der Sammlung von Zuweisungen enzymatischer Funktionen aus den bioinformatischen Datenbanken des NCBI, KEGG, PEDANT, BRENDA, AMENDA und SwissProt (Kapitel 1.9) und deren zugeteilten Konfidenzen (Tabelle 1), dienen dem EnzymeDetector auch eigene Algorithmen zur Genannotation. Zum Beispiel offeriert ein bioinformatischer Sequenzvergleich (BLAST, Kapitel 2.2) der zu annotierenden Gene zu einer Sammlung von Sequenzen bereits bekannter Funktionen (UniProt, Kapitel 1.9) weitere Aspekte für die Wahrscheinlichkeit entsprechender Zuweisungen. Zur Bewertung dieser Vorhersagemethode dienen der Erwartungswert des BLASTs (Kapitel 2.2) als direktes Gütemerkmal und verschiedene Schwellenwerte zur Filterung der Ergebnisse. Dementsprechend müssen für die maximale Konfidenz von 8 Bewertungspunkten in der annotierten Vergleichsdatenbank mindestens 5 verschiedene Sequenzen bei gleichwertiger Ähnlichkeit für die selbe enzymatische Funktion codieren und einen Erwartungswert kleiner als 10^{-120} erhalten. Infolgedessen werden zufällige Sequenzübereinstimmungen nahezu ausgeschlossen und eine entsprechend vielseitige Beschreibung der Enzymfunktion bei mindestens 5 Sequenzen verschiedener Herkunft und somit zumeist verschiedener Organismen vorausgesetzt. Weitere Abstufungen der Trefferqualität erfolgen nach Bereichseinordnungen der erhaltenen Erwartungswerte, wobei dieser 10^{-3} nicht übersteigt (Tabelle 1). Zusätzliche Ergebniswerte des BLASTs wie die Positionsangaben der übereinstimmenden Sequenzabschnitte und die absolute Identität als Maß für die Deckungsgleichheit der gegenübergestellten Aminosäuresequenzen werden nicht mit in die Ergebnistabellen übernommen und können keiner weiterführenden Auswertung dienen. Sehr häufig codieren allerdings Gensequenzen für zwei oder mehr voneinander verschiedene enzymatische Funktionen. Diese sogenannten bi-, bzw., multifunktionellen Enzyme übernehmen je nach induzierten Bedarf des Metabolismus eines Organismus und infolgedessen der differenzierten Translation der Sequenzen die Katalyse verschiedener Reaktionen. [34,35] Dementsprechend können den einzelnen Abschnitten der Gensequenz unterschiedliche Enzymfunktionen zugeordnet werden. Voraussetzung hierfür ist allerdings die genaue Bekanntheit der entsprechenden Sequenzabschnitte und deren Positionen zur Einschätzung der Lage auf dem Gen. Dies ermöglicht der EnzymeDetector bisher nicht.

Als weiterer wichtiger Aspekt für die Differenzierung von Enzymannotationen dient der Vergleich mit sogenannten Sequenzmustern. In entsprechenden Analysen bekannter Enzymsequenzen werden zum Beispiel bestimmte Wiederholungen und Kombinationen der Aminosäuresequenz erkannt und mit entsprechenden Mustern abgebildet. Diese Datenbank namens BrEPS (Kapitel 1.9) liefert mit der bestimmten Korrektheit (*Correctness*) der Mustererkennung von Enzymfunktionen ein direktes Gütemerkmal. [36] Demgemäß und infolge der Anzahl der erkannten Enzymklassen durch ein einzelnes Muster, also der

Spezifität, erfolgt in der Ergebnismenge des EnzymeDetectors eine Einordnung in Einzelkonfidenzen von 1-10.

Tabelle 1: Scoring-Schema des EnzymeDetector im Original. Die Tabelle listet die einzelnen Bewertungen des EnzymeDetector auf. Der maximal erreichbare Score einer Quelle entspricht dem Konfidenzwert. Dabei werden die einzelnen Abstufungen der BLAST und BrEPS Konfidenzwerte gesondert aufgeführt.

Quelle	Konfidenzwert
NCBI, KEGG, PEDANT	3
BLAST	E-Value < 10 ⁻¹²⁰ : 8 E-Value < 10 ⁻⁸⁰ : 6 E-Value < 10 ⁻⁴⁰ : 4 E-Value > 10 ⁻⁴⁰ : 2
BrEPS	1-10, nach <i>Correctness</i> des Sequenzmusters
SwissProt	50
AMENDA	25
BRENDA	50

Die Einstufung von Güteklassen der entsprechenden Annotationen erfolgt ausschließlich bei der Vorhersagemethode mittels BLAST und BrEPS. Die Konfidenz der übrigen Quellen wird entweder bei Übereinstimmung zugewiesen oder nicht. Demzufolge erhalten Zusammenhänge von Enzymen und Organismen aus der automatisierten Textmining Datenbank AMENDA (Kapitel 1.9), ohne jegliche Beachtung der Art der Kookkurrenz in denen sie stehen, eine maximale Konfidenz von 25. Ebenso ungeachtet der Informationen zu Organismusstämmen und Verlinkungen zu Sequenzen der Annotationen von Enzymen der BRENDA (Kapitel 1.9) erhalten entsprechende Zuordnungen grundsätzlich eine maximale Bewertung von 50 genauso wie die Zuweisungen der SwissProt-Datenbank. Damit ist gegeben, dass hochwertige Informationen dieser manuell geprüften Datenbanken stets höher bewertet werden, als die Summe aller Einzelkonfidenzen der automatisierten Quellen. Demzufolge ist sichergestellt, dass Informationen dieser Güte nicht überstimmt werden können. Dies wiederum verdeutlicht die Pflicht zur sensiblen Vergabe solcher Zuordnungen nur auf korrekten und fundamentierten Fakten basierend. Demgemäß ist die strikte Höchstbewertung ebenfalls zu überdenken, wie die zwangsweise angelegte Interpretation von enzymatischen Funktionen in jegliche Sequenzen, die dem BLAST-Algorithmus des EnzymeDetector vorliegen. Denn demgegenüber steht die Tatsache einer statistischen Analyse von Annotationsprojekten, die besagt, dass zumeist etwa 30% eines Genoms für enzymatische Funktionen codieren. [23] Nebenbei erwähnt, können diese 30% des Genoms bei ausgewählt untersuchten Organismen nur erreicht werden, wenn die Informationen mehrerer Annotationsquellen einfließen. [23] Dies verdeutlicht wiederholt die Wichtigkeit des EnzymeDetector für die Anwendung und Forschungen auf Basis von Genannotationen zum Beispiel im Bereich der Systembiologie.

Die Ergebnisse der letztendlich resultierenden Annotationen aller im EnzymeDetector untersuchten Organismen können einerseits über eine Weboberfläche (<http://edbs.tu-bs.de>) öffentlich zugänglich in Augenschein genommen und andererseits über organismusspezifische Tabellen einer institutsinternen MySQL-Datenbank ausgewertet werden.

Die Wertigkeit der Konfidenz-Aussage des EnzymeDetector ist direkt abhängig von der Anzahl einfließender Annotationsquellen. Die Einbindung zusätzlicher Datenbanken und Methoden zur Funktionsvorhersage ist demzufolge ein wichtiger Aspekt der Weiterentwicklung des Programms. Für die Abbildung der Annotationen verschiedener Quelldatenbanken werden sogenannte *Geneidentifizier* des NCBI herangezogen. Diese sind allerdings sequenzeindeutig ohne direkten Bezug zum Genom von Organismen, in denen sie vorkommen können und unterliegen einem fortwährenden Wandel im Laufe von Datenbankupdates des NCBI (Kapitel 1.9) [25]. Häufig können somit auf dieser Grundlage Vorhersageinformationen diverser Datenbanken dem Genom nicht eindeutig zugeordnet werden, wodurch diese in der Ergebnismenge des EnzymeDetector verworfen werden.

Des Weiteren ist die Anzahl annotierter Genome diverser Spezies und Organismenstämme ein entscheidender Faktor für die Anwendung des Programms von Wissenschaftlern bei der Untersuchung verschiedenster Bakterien und der Analyse von Vergleichen des jeweiligen Metabolismus. Der EnzymeDetector auf dem Stand von 08/2012 beschreibt und sammelt Annotationen der zuvor aufgeführten Quellen zu 189 Organismen, während als zeitkritischer Schritt die BLAST-Analyse ausgemacht wird.

Auf den Nutzen und die Wichtigkeit des Programms EnzymeDetector für das Wissenschaftsgebiet der Systembiologie und der Enzymannotation als essentiellern Ausgangspunkt für die Erstellung metabolischer Netzwerke und Modelle untersuchter Organismen ist in diesem Kapitel mehrfach hingewiesen. Nichtsdestotrotz sind einige Erweiterungen, Optimierungen und Korrekturen unumgänglich und Gegenstand der vorliegenden Arbeit. (Kapitel 2.6 und Kapitel 3.1)

1.7 Die Domäne der Archaea

Allgemeines

Die Archaeen bilden neben den Bakterien und Eukaryoten eine der drei Domänen der systematischen Biologie. Sie gehören zu den einzelligen Lebewesen und besitzen ein in sich geschlossenes DNA-Molekül als Kernäquivalent. [37,38] Folglich zählen sie zu den Prokaryoten. Diverse strukturelle, physiologische, biochemische und weiterhin genetische Eigenschaften charakterisieren diese Organismen und begründen die separierte Stellung in der Phylogenie genauso wie teilweise beschriebene Eigenheiten des Metabolismus dieser Vertreter. Besonders deutlich wird dies allerdings durch die Sequenzunterschiede der RNA der kleinen ribosomalen Untereinheit, der 16S rRNA. Im Vergleich mit den Bakterien weisen sie geringere Unterschiede bei der Gegenüberstellung eukaryotischer Sequenzen auf. [38] Dieser Sequenzmarker ist Bestandteil jeder lebenden Zelle und übernimmt in allen Organismen dieselbe Funktion bei der Translation von Genen. Da die ribosomalen Strukturen ebenso wie die DNA besonderen Schutzmechanismen obliegen und dazu nur selten dem Gentransfer zwischen Organismen unterliegen, wird ein geringer Mutationsdruck auf diesen Sequenzen erwartet, sodass die evolutive Entwicklungsgeschichte durch ihre Konstanz über alle Lebewesen hinweg sehr gut nachvollzogen werden kann. [39,40]

Eigenschaften der Archaeen

Wie bereits erwähnt, sind einerseits einige molekularbiologische Mechanismen zum Beispiel der Translation und Transkription der Archaeen sehr ähnlich zu der Domäne der Eukaryota.

Auf der anderen Seite gibt es deutliche Parallelen zu den Bakterien, wie den fehlenden Zellkern, die Zellgröße von etwa 1-20 µm, die Art der Zellteilung mit den beteiligten Proteinen und relativ basal aufgebaute Fortbewegungsorgane, die sogenannten Flagellen. Das Genom ist ebenso wie bei den Bakterien in Operons organisiert. Das bedeutet, dass funktionelle Einheiten auf dem Genom direkt hintereinander codiert sind und aus der Promotor-/Operatoreinheit und eventuell mehreren regulatorischen Proteinen zur Expressionsregulierung bestehen. [14] Dazu codieren weiterhin mehrere Strukturgene für Proteine, die letztendlich für die Funktionen des Operons verantwortlich sind und im funktionellen Zusammenhang stehen. [41] Ein genetisch wichtiger Aspekt ist ferner die Möglichkeit, dass Archaeengenome auch Plasmide tragen. Plasmide sind für normal ringförmige und doppelstrangige DNA-Moleküle, die in Bakterien und in Archaeen vorkommen können, aber nicht zum eigentlichen Kernäquivalent zu zählen sind. Sie liegen folglich extrachromosomal vor und besitzen die Fähigkeit, sich autonom zu replizieren. [33] Dennoch besitzen Archaeen individuelle Eigenschaften, die in dieser Form in keiner anderen Domäne zu finden sind. Dazu zählt beispielsweise der Aufbau der Zellwand und Zellmembran mit deutlichen Unterschieden im Aufbau zu den Bakterien und Eukaryoten. Die Zellwände der Archaeen enthalten im Grundaufbau ein Polysaccharid mit vorrangig kurzen, gebundenen Aminosäureketten, das Pseudopeptidoglycan. Über die gesamte Domäne hinweg sind vielfältige weitere Anpassungen und Strukturen im Aufbau der Zellwand und Zellmembran zu beobachten. [38] Demzufolge verzichten manche Archaeen der Spezies *Thermoplasma* auf eine Zellwand [42], während andere hochkomplexe Strukturen aufbauen. Biochemisch unterscheidet sich die Zellmembran durch die Art der zu findenden Bindungen. Bei Bakterien und Eukaryoten sind Fettsäuren über eine einfache Esterbindung an Glycerolmoleküle gebunden. Währenddessen binden Archaeen hier stattdessen verzweigte Isopreneinheiten über Glyceroldiether oder Glyceroltetraether. Die Isoprene sind ungesättigte Kohlenwasserstoffe, die in der Membran die hydrophoben Anteile der Lipide bilden. Derart stabilisierte Strukturen sind vornehmlich bei hyperthermophilen Archaeen [43] aus extremsten Temperaturbereichen zu finden und können ebenfalls Vertretern in stark sauren Habitaten den entsprechenden Schutz zur Umwelt bieten [44]. Die meisten Archaeen sind mobil und können sich mit Hilfe ihrer Flagellen in Relation zu den Bakterien recht schnell fortbewegen. [45]

Habitate der Archaeen

Die meisten in der Literatur beschriebenen Archaeen sind extremophil. Das bedeutet, sie leben in Habitaten sehr niedriger oder hoher Temperaturen (psychrophil, thermophil), sehr niedriger oder hoher pH-Werte (acidophil bzw. alkaliphil) und teils sehr hohen Salzkonzentrationen (halophil). Dies erfordert unweigerlich diverse Anpassungen an diese extremen Biotope. [38]

Die Archaeen machen große Anteile der Lebewesen im Meer, aber auch in Boden- und Süßwasserbiotopen aus. Hyperthermophile Archaeen sind häufig in

marinen und terrestrischen vulkanischen Gebieten in sogenannten heißen oder Schwefelquellen die einzigen Organismen, die unter den dort vorherrschenden Bedingungen überleben können. Halophile Vertreter bevorzugen hohe Salzgehalte der Meere und extremste Salzspots, wie sie zum Beispiel im Toten Meer zu finden sind. [38]

Außerdem leben einige Archaeen unter anoxischen Bedingungen. Diese sauerstofffreien Habitate setzen eine Anpassung des Metabolismus voraus und Vertreter sind im Schlamm und Schlick von Süßwasser, im Meer und im Boden zu finden. Ebenfalls werden Archaeenspezies auch als Symbionten im Darmtrakt von Tieren und Menschen lebend beschrieben. [38]

Die Extremophilie der Archaeen sorgt immer wieder für Probleme bei der Kultivierung im Labor, sodass beispielsweise bisher nur sehr wenige kälteliebende Spezies als Reinkultur verfügbar sind. Dennoch wird allein durch die breitgefächerten Habitate deutlich, dass Archaeen eine entscheidende Rolle im gesamten Ökosystem aller Lebewesen einnehmen. [38]

Stoffwechsel der Archaeen

Die Vertreter der Archaeen können teils sowohl einen autotrophen Stoffwechsel zur Assimilation von Kohlenstoff aus anorganischem Kohlenstoffdioxid aufweisen, als auch organische Verbindungen als Kohlenstoffquelle verwenden (heterotroph). [38]

Weiterhin zeichnet viele Archaeen ein anaerober Stoffwechsel ohne Sauerstoff aus. Für viele dieser Vertreter ist Sauerstoff sogar toxisch. Eine Besonderheit des Metabolismus der Archaeen stellt die Methanogenese dar. Hierbei wird anaerob Biomasse abgebaut und mit Hilfe von Kohlenstoffdioxid und Wasserstoff Methan gebildet. Währenddessen wird Energie freigesetzt und steht den Organismen zur Verfügung. Die methanogenen Archaeen sind charakterisiert durch eine Reihe von einzigartigen Cofaktoren, wie dem Coenzym F₄₂₀ oder Methanofuran, die auf ihrem Genom codiert sind. [14]

Die meisten Archaeen aus hohen Temperaturbereichen sind Anaerobier und mit einem chemoorganotrophen (Energiegewinnung aus der chemischen Umsetzung von organischen Substanzen) oder chemolithotrophen (Energiegewinnung aus der chemischen Umsetzung von anorganischen Substanzen) Stoffwechsel zu beschreiben. Oft spielt bei letzterem der Abbau von Schwefelverbindungen eine entscheidende Rolle. [38] Insofern wird Energie durch die Reduzierung von Schwefel freigesetzt. Dies ist auch bei aeroben Arten wie *Acidianus ambivalens* beschrieben, welche den Schwefel mit spezifischen Enzymen zur Energiegewinnung oxidieren. [46] Oftmals sind halophile Vertreter mit einem aeroben Metabolismus ausgestattet, während sie ihre Energie aus organischen Verbindungen gewinnen. [47–49]

Morphologie der Archaeen

Die Archaeen besitzen wie bereits angesprochen Flagellen zur Fortbewegung und oft auch Pili, ein fadenförmiger Zellfortsatz bestehend aus verschiedenen Proteinen [14], zur Anheftung an entsprechende Oberflächen. Diverse Zellformen von Kokken, Stäbchen, Spirillen, scheiben- oder filamentartig und darüber hinaus auch quadratisch (zum Beispiel *Haloquadratum walsbyi* [50]) sind bei diesen Organismen beschrieben. [38]

Systematik der Archaeen

Die systematische Einteilung der Archaeen war lange Zeit im Wandel. Allerdings kann die phylogenetische Phylumeinteilung der thermophilen Crenarchaeota, der halophilen Euarchaeota inklusive der Untergruppe der methanogenen Vertreter und einer Sondergruppe der Kor- und Thaumarchaeota mit heutigen 16S rRNA-Analysen bestätigt werden. [38,51,52] Eine weitere Sondergruppe der Euarchaeota, der bisweilen die anerkannte phylogenetische Validierung auf Phylumbene fehlt, stellen die oft extrem acidophilen Archaeen der Thermoplasmata dar. Sie können bei pH-Werten nahe 0 überleben und untergliedern sich in die Familien der Ferroplasmaceae, Picrophilaceae und Thermoplasmataceae. [38,53]

Forschung über Archaeen

Die Diversität der Archaeen, ihre spezifischen Anpassungen an extreme Habitate und ihre tragende ökologische Rolle machen diese Organismen interessant für aktuelle Forschungen und Untersuchungen der Biotechnologie, aber auch der Grundlagenforschung und Systembiologie. Demzufolge werden Archaeen zur Biogasgewinnung und deren Optimierung intensiv untersucht. [54] Die Medizin versucht durch die Verwendung von archaealen Zellwandbestandteilen, neue Träger für Impfstoffe herzustellen (S-Layer). [55–57] Die Biochemie kann beispielsweise neue hitzeresistente Enzyme beschreiben und isolieren. Diese können wiederum in der Laborforschung der Genetik in Zukunft tragende Rollen einnehmen. [58–60]

Der Systembiologie stehen einerseits besondere enzymatische Funktionen unter harschen Bedingungen mit archaeentypischer breiter Substratspezifität zur Forschung offen und besondere, aber auch äußerst diverse metabolische Abläufe hält diese Domäne bereit. Aus diesem Grund werden die Archaeen in dieser Arbeit als Organismengruppe genauer untersucht.

1.8 Die Familie der Rhodobacteraceae

Allgemeines und Eigenschaften der Rhodobacteraceen

Die Familie der Rhodobacteraceae stellt eine Gruppe gram-negativer alpha-Proteobakterien dar. Vertreter dieser Organismen gehören zu den häufigsten Prokaryoten in marinen Habitaten und Ökosystemen. Sie spielen für die globalen Stoffkreisläufe der Meere eine essentielle Rolle. Dies ist einerseits durch ihre physiologische Vielseitigkeit und andererseits durch die gruppenintern hohe Diversität der Eigenschaften und metabolischen Abläufe zu begründen. [61–63] Weitere wichtige Aspekte sind die in der Familie weit verbreiteten Möglichkeiten zur Biofilmbildung [64] und das Leben in Symbiose mit Arten der Algen. [65] Die Aggregation vielzähliger Individuen birgt den Vorteil, durch die fortwährende Adsorption der zu verbrauchenden Verbindungen, einer besseren Nährstoffversorgung an der Oberfläche solcher Biofilme. Ebenso kann eine Schutzfunktion gegenüber toxischen Substanzen durch die entstehende Diffusionsbarriere und dem physiologischen Zustand der adhärenierten Bakterien aufgebaut werden. [66] Die Assoziation mit den Algen entspricht von Spezies zu Spezies mehr oder weniger einer echten Symbiose. Während einige Vertreter diese Verbindung zum beiderseitigen Nutzen pflegen, induzieren andere nur die

Produktion von benötigten Kohlenstoffverbindungen durch die Alge. [67,68] Hingegen ist nachgewiesen, dass beispielsweise *Dinoroseobacter shibae* dem symbiontischen Partner wichtige wachstumslimitierende Vitamine (B1, B12) bereitstellt. [69]

Habitate der Rhodobacteraceen

Die meisten Arten sind im Meer- und Süßwasser anzutreffen. Allerdings zählen zu der Familie auch Vertreter, die terrestrisch im Boden, im Ab-/Brackwasser und im Schlamm oder Sediment von Gewässern leben. [63]

Stoffwechsel der Rhodobacteraceen

Der Sauerstoffmetabolismus der Rhodobacteraceen ist zumeist aerob beschrieben, während einige denitrifizierende Vertreter wie beispielsweise *Paracoccus denitrificans* und *Pseudovibrio denitrificans* allerdings ebenso fakultativ anaerob Wachstum unter sauerstofffreien Bedingungen nachweisen. [70,71] Einigen Vertretern dienen Licht als Energie- und organische Stoffe als Kohlenstoffquelle. Diese photoheterotrophen Organismen besitzen die Fähigkeit einer anoxygenen Photosynthese, ohne Sauerstoff zu produzieren. [72] Diesbezüglich besitzen sie auch Mechanismen und Strukturen, um sich dem Licht zugewandt auszurichten. [73] Manchen Arten ist es weiterhin möglich, mit Hilfe der CO₂-Fixierung als Kohlenstoffquelle photoautotroph zu wachsen. [74] Dabei werden teils bestimmte Schwefelverbindungen wie Sulfide oder Thiosulfate als Elektronendonatoren genutzt. [75] Bei einigen anaeroben Vertretern sind Fermentationsstoffwechselwege beschrieben. Hierbei werden Zucker zu Säuren, Gasen oder Alkohol verstoffwechselt, um ohne Sauerstoff durch enzymatische Katalysen Energie zu gewinnen. [2,76]

Morphologie der Rhodobacteraceen

Bei den Rhodobacteraceen sind Kokken und ei- oder stäbchenförmige Zellen beschrieben. Viele Vertreter sind begeißelt und infolgedessen beweglich in ihrem Medium. [61,62]

Systematik der Rhodobacteraceen

Die phylogenetische Einteilung ist fortwährend im Wandel. Demzufolge werden oft Stämme und Spezies reklassifiziert, nachdem sie näheren Analysen unterlaufen sind. [77] Des Weiteren wird in einschlägiger Literatur immer häufiger die Frage aufgeworfen, ob die Familie der Rhodobacteraceae fortwährend als Klade zu bezeichnen ist, oder es sich hier auf Grund entscheidender Differenzen in mikrobiellen Eigenschaften, der Physiologie und im Metabolismus treffender um Organismusgruppierungen handelt. [61,67,78]

Forschung über Rhodobacteraceen

Diese sehr weit in marinen Gewässern verbreitete und entsprechend erfolgreiche Familie von Bakterien weckt das Interesse der aktuellen Forschung. Gegenüber dieser hohen ökologischen Relevanz, sind noch relativ wenige metabolische und genetische Zusammenhänge dieser Gruppe beschrieben. Ebenfalls phylogenetische Aspekte führen immer noch zu offenen Fragen, ob eine

derartige Diversität einer Familie überhaupt tragbar ist oder hier weitere Gruppierungen vorgenommen werden müssen. Das Verständnis der evolutionären, genetischen und physiologischen Prinzipien dieser Familie ist hochbedeutend für die Systembiologie und kann charakteristische Modellorganismen für erfolgreich an verschiedene Lebensräume angepasste Vertreter mit entsprechenden metabolischen Abläufen und sowohl genetischen, als auch strukturellen Anpassungen hervorbringen. [79–81] Folglich werden auch Enzymfunktionen und der Metabolismus dieser Organismengruppe in dieser Arbeit untersucht.

1.9 Die Sequenz-, Annotations-, Stoffwechselweg- und Metadatenquellen

Die Bioinformatik und die Entwicklung diverser Methoden zur Auswertung und Weiterverarbeitung experimenteller Daten führte im Laufe der Zeit zu extremsten Datenmengen, die gespeichert und abrufbar Zusammenhänge erkennen lassen. Die grundlegendsten Daten von Sequenzierungsprojekten zur Analyse der Genome untersuchter Organismen werden in den Sequenzdatenbanken in Form relationaler Konzepte [82] und simplen Textdatendateien organisiert. In weiteren Datenbanken werden Annotationsvorhersagen enzymatischer Funktionen durch entsprechende Methoden zur Verfügung gestellt. Wiederum weitere Datensammlungen beschreiben Stoffwechselwege, indem sie Metabolite durch enzymatische Reaktionen in Verbindung stellen und im Gesamtkontext des Metabolismus von Organismen interpretieren. Dazu dienen ebenfalls, insofern verfügbar, experimentelle Nachweise als Beleg dieser Reaktionsketten und Mechanismen.

NCBI

Eine Vielzahl von Datenbanken biologischer Informationen und Sequenzierungsprojekten größtenteils der GenBank [83] stellt das *National Center for Biotechnology Information (NCBI)* [26,84] als Teil des *National Institutes of Health (NIH)* aus den USA zur Verfügung. Über verschiedene öffentlich im Internet zugängliche Schnittstellen werden DNA-, RNA- und Proteinsequenzdaten (*RefSeq* [25]) mit den Informationen zu den jeweiligen Sequenzierungsprojekten (*BioProject*, *BioSample*) zugänglich gemacht. Dabei werden den meisten Genen über eine automatisierte Annotationspipeline auf der Basis von Sequenzähnlichkeiten, Analysen kooperierender Arbeitsgruppen und teils durch manuelle, experimentell belegte Informationen Funktionsvorhersagen auch im Hinblick auf Enzymklassen zugeordnet.

Weiterhin wird durch die Sammlung phylogenetischer Informationen aus publizierter Literatur, verschiedenen Datenbanken und den Hinweisen von Experten der Phylogenie ein umfassender Taxonomischer Baum über alle Lebewesen für eigene Analysen und Daten zu bestimmten Spezies angeboten. [84] Zu guter Letzt wird neben diverser Standardsoftware der Bioinformatik eine umfassende Sammlung an Literaturverlinkungen medizinischer Fachartikel (*PubMed*) zur eigenen Recherche frei zugänglich angeboten. Eine Standardsoftware ist der Alignment-Algorithmus namens **Basic Local Alignment Search Tool** (BLAST, Kapitel 2.2). Das Programm wird dazu verwendet, DNA- oder Proteinsequenzen im Hinblick auf die Ähnlichkeit und Identität mit einer Sequenzdatenbank zu vergleichen. In einem heuristischen Verfahren werden

erst kleine Sequenzabschnitte mit hoher Übereinstimmung gesucht und anschließend, insofern möglich, entsprechend erweitert. Mit diesem Verfahren können derartige Alignments zur Findung homologer Sequenzen in kurzer Zeit erstellt werden und somit Datenbanken nach Genen gleicher Funktion abgefragt werden. Zur Einschätzung der Qualität werden diverse Ergebnisparameter vermittelt. Demgemäß liegt mit dem sogenannten Score eine quantitative Bewertung der Ähnlichkeit der verglichenen Sequenzen vor. Je höher der Score ausfällt, desto besser die Identität und Ähnlichkeit der Sequenzen. Der *Expected (E)-Value* ist ein Maß für die Wahrscheinlichkeit, dass ein gleichwertiger Treffer per Zufall in der durchsuchten Datenbank gefunden wird. Damit stellt dieser ein Maß für die Signifikanz der Suchtreffer dar und ist umso besser einzustufen, umso mehr der Wert gegen Null geht. Mit Hilfe der weiteren Angaben, wie der Identität der Sequenzen, der Alignmentlänge, an der die Sequenzen aneinander durch gleiche oder sehr ähnliche Aminosäuren ausgerichtet werden können, und der Informationen der Treffersequenz selbst, können Schlussfolgerungen über homologe Funktionen getroffen werden.

KEGG

Der Bereitstellung von Stoffwechselwegkarten hat sich die integrierte Datenbank der *Kyoto Encyclopedia of Genes and Genomes* (**KEGG**) [22,24] hauptsächlich verschrieben. Darüber hinaus werden aber auch genomische und chemische Daten angeboten und die Ergebnisse zweier Enzymannotationsalgorithmen publiziert [22,85]. Dazu werden zum Beispiel homologe Sequenzen aller bekannten Organismen zusammen gruppiert und Funktionen abgeleitet (Orthologie-Cluster). Die frei zugängliche Datenbank verfolgt das Ziel, durch die Bereitstellung dieser bioinformatischen Daten aller bekannten Organismen die computergestützte Vorhersage komplexer Zellprozesse zu forcieren und genetische, bzw. molekulare Informationen in Zusammenhang zu bringen.

PEDANT

Die Datenbank des *Protein Extraction, Description and ANalysis Tool* [27] bietet durch die Kombination bioinformatischer Tools eine automatisierte Analyse zur Funktionsvorhersage verschiedener Organismussequenzen an. Eingesetzt werden neben einem sehr stringenten BLAST (Kapitel 2.2) gegen eine nicht redundante Sequenzdatenbank bekannter Funktionen, auch Motivsuchen mittels Proteinfamilien (Pfam) [86] und der Vergleich funktionsentscheidender Abschnitte und Domänen von Proteinsequenzen (BLOCKS [87] und PROSITE [88] Datenbank).

PATRIC

Eine weitere Datenquelle bietet sich mit dem *Pathosystems Resource Integration Center*, **PATRIC** [89] an. Sie nimmt sich generell der Forschung an bakteriellen Infektionskrankheiten an und erlaubt weiterhin durch ihre eigene Annotationspipeline mit Hilfe des automatischen Annotationssystems von Mikroorganismen „*Rapid Annotation using Subsystem Technology*“ (RAST) [90] die Zuordnung von enzymatischen Funktionen zu den Genen von über 5000 bakteriellen und 200 archaealen Genomen. Neben diesen Funktionsvorhersagen bietet PATRIC eine Plattform für biomedizinische Forscher zur Suche

spezifischer Organismendaten verlinkt zu weiteren Datenbanken und Programme zur komparativen Genomik, das heißt dem Vergleich sequenzierter Genome zu Referenzorganismen und der Analyse der Ergebnisse, an.

UniProt

Auch die *Universal Protein Resource* (**UniProt**) [32] sammelt neben vielen bioinformatischen Informationen zu Proteomen und Proteinen die Zuordnungen von Enzymfunktionen zu Aminosäuresequenzen. Die Sammlung insgesamt setzt sich zusammen aus der UniProt *Knowledgebase* (UniProtKB), den gruppierten Sequenzsets UniProt *Reference Clusters* (UniRef) und einem nicht redundanten UniProt Archiv (UniParc). Eine Teilmenge der UniProtKB namens **SwissProt** vermerkt besonders verlässliche Enzymannotationen. Diese Einträge sind im besten Fall experimentell nachgewiesen oder zumindest in der Literatur publiziert und verwiesen. Auch werden geprüfte automatisiert erstellte Enzymannotationen nach und nach dieser Datenbank zugeordnet. Dem gegenüber sind Annotationen, die rein automatisiert durch Sequenzvergleiche, folglich Homologien und andere bioinformatische Methoden generiert werden, in der *Translated EMBL Nucleotide Sequence Data Library* (TrEMBL) Datenbank zu finden.

Pfam

Eine weitere Möglichkeit, Proteine zu gruppieren, gestattet die Unterteilung in Proteinfamilien. Die Tatsache, dass das Vorkommen einzelner Proteindomänen in verschiedene Architekturen im Zusammenhang mit der Proteinfunktion steht, macht sich die **Pfam**-Datenbank des European Bioinformatics Institute [86] zunutze. Die frei über einen Webservice zugängliche Datenbank stellt für alle bekannten Proteinfunktionen die charakteristischen Proteindomänen zusammen mit sequenzbeschreibenden Profilen zur Verfügung (HMM, Kapitel 2.2) und kategorisiert die Proteine entsprechend der Mustererkennung.

BREPS

Ebenfalls über Sequenzmuster (*Pattern*) ordnet die **Braunschweig Enzyme Pattern Search** [36] Proteinsequenzen Funktionen zu. Dazu werden mit Hilfe von BLAST (Kapitel 2.2) die Sequenzen der SwissProt-Datenbank gruppiert und die entstandenen Einheiten aligniert. Das bedeutet, dass diese homologen Sequenzen an identischen und sehr ähnlichen Abschnitten ausgerichtet werden, um einerseits variable Bereiche ohne entscheidenden Einfluss auf die Proteinfunktion zu erkennen und andererseits typische Muster extrahieren zu können. Die so erhaltenen Sequenzmuster stehen anschließend nach der Enzymklasse und der Güte (*Correctness*) kategorisiert zur Verfügung, um Funktionen unbekannter Sequenzen zu erkennen. Die Enzymklasse wird durch die annotierten EC-Nummern in SwissProt der betrachteten Sequenzgruppe zugeordnet, sodass durchaus Muster für mehrere enzymatische Funktionen stehen können. Die Güte wird entsprechend dem Wiederfinden der korrekten Enzyme, für die das Sequenzmuster codiert, aus der gesamten SwissProt-Datenbank und der Spezifität bestimmt.

BRENDA

Zu den weltweit anerkanntesten und umfangreichsten Biochemie Datenbanken zählt die **BR**aunschweig **EN**zyme **DA**tabase [20]. Das online verfügbare Enzyminformationssystem wird einerseits dazu genutzt, um sowohl einen detaillierten Überblick zu funktionellen Eigenschaften von Enzymen zu erhalten, als auch biochemische Zusammenhänge in Bezug auf Stoffwechselwege und Verknüpfungen zu Organismen und Funktionen im Hinblick auf wissenschaftliche Publikationen zu erkennen.

Zu diesem Zweck bietet das Webinterface diverse Suchmasken für schnelle und erweiterte Abfragen und die namensbasierte oder chemische Struktursuche zur Ermittlung von Verbindungen in Interaktion mit Enzymklassen. Dazu zählen Substrate, Cofaktoren und Inhibitoren von Enzymen genauso wie die zugeordneten enzymkatalysierten Reaktionen. Weiterhin dienen ein kategorisierter Enzymbaum und eine taxonomische Ordnung annotierter Enzyme nach den Organismen, in denen sie vorkommen, der schnellen Auffindung enzymespezifischer Daten und Informationen, die von Biochemikern, Biologen und Chemikern manuell aus wissenschaftlicher Literatur extrahiert werden oder aus entsprechend verlässlichen automatisierten Vorhersagen stammen.

Dabei ist eine große Sammlung synonyme Bezeichnungen der Liganden, Substrate und Enzyme von immenser Bedeutung, um einerseits Daten zuverlässig auffindig zu machen und andererseits Informationen verschiedener Quellen zu aggregieren.

Eine der ersten gewebespezifischen Ontologien im Bereich der Biowissenschaften, die Verknüpfungen der Enzyme zu den taxonomischen Gruppen der Metazoa, Plantae, Fungi und Protozoen, in denen sie nachgewiesen sind, herstellt, ist ebenfalls Gegenstand dieser Datenbank. Dementsprechend werden Bezeichnungen für Gewebe, Organe, anatomische Strukturen, Pflanzenteile, Zellkulturen, Zelltypen und Zelllinien der Organismen den enzymatischen Funktionen in der **BRENDA Tissue Ontology** Datenbank zugeordnet. [91]

AMENDA

Ein Teil der automatisierten Vorhersagen von Enzymklassen verschiedener Organismen der BRENDA stammt aus der zusätzlichen Datenbank namens **AMENDA**. [92] Demzufolge sind in der experimentellen Literatur belegte, sequenzbasierte und in Folge dessen einem Genlocus eines Organismus zugeordnete Annotationen von denen zu unterscheiden, die durch ein automatisiertes Textmining (***A**utomatic **M**ining of **EN**zyme **D**ata*) generiert sind. Diese genunabhängigen Annotationen geben entscheidende Hinweise auf Enzymfunktionen im Stoffwechsel eines untersuchten Organismus – ein entscheidendes Kriterium für Modellierer, welches allerdings einer qualitativen Einordnung bedarf. Die Grundlage für das Textmining schafft die Literaturdatenbank PubMed [84]. Um entsprechende Verknüpfungen von Organismen und enzymatischen Funktionen zu erhalten, werden alle Titel und Kurzbeschreibungen der verlinkten wissenschaftlichen Literatur nach Textbausteinen und bestimmten Begriffen mehr oder weniger zusammenstehend automatisch durchsucht und die Ergebnisse gespeichert. Eine Aufbereitung und Bewertung dieser Informationen erfolgt durch die Einordnung in verschiedene

Qualitätskategorien. Dabei wird eine Wertung vorgenommen, inwieweit die Bezeichnungen von Enzymen direkt als Konkurrenz mit den Organismennamen und -stämmen auftreten und differenziert, ob eine positive oder negierte Aussage getroffen wird. Demgemäß erhält beispielsweise eine enzymatische Funktion unmittelbar gefolgt von einer Spezies im Titel der Publikation die Höchstbewertung. Entsprechende Abstufungen werden nach Ort (Titel, Kurzbeschreibung) und Nähe der Okkurrenz (Wort, Satz, Absatz) vorgenommen und so auf wahrscheinliche Zusammenhänge hingewiesen.

MetaCyc

Eine geprüfte und organisierte Zusammenstellung nicht redundanter und experimentell nachvollzogener metabolischer Stoffwechselwege stellt **MetaCyc** [21] dar. Dabei werden experimentelle Hinweise auf zusammenhängende Reaktionsketten aus allen Domänen der systematischen Biologie berücksichtigt, um metabolische Stoffwechselwege vorherzusagen, diese mit ihren Metaboliten in Zusammenhang zu bringen und entsprechende beteiligte und benötigte Enzyme zu verknüpfen. Dabei wird aus den beschriebenen Organismen der experimentellen Grundlage jeden Weges eine sogenannte *taxonomic range* abgeleitet, um zu verdeutlichen, wo diese Abläufe bisher beobachtet wurden und in welchen Gruppen der Taxonomie äquivalente Vorgänge zu erwarten sind.

Weiterhin wird dem Benutzer angeboten, untersuchte Stoffwechselwege als Bildcollage der Substrate, der Produkte, der Reaktionsrichtungen und der beteiligten Enzymklassen zu exportieren. Dazu dient eine sogenannte *reaction-layout*-Datenbank, die die Reaktionsketten zur Umwandlungen der Metabolite in ein *in-silico* verarbeitbares Format überführt. Hierbei werden Cofaktoren, Enzyme und beteiligte Reaktionspartner alle gleichermaßen registriert.

Die verzeichneten Stoffwechselwege müssen keinesfalls linear aufgebaut sein. Zyklische und verzweigte Abläufe und Prozesse spielen ebenso eine Rolle, wie die Erkennung und Gruppierung alternativer Reaktionen, die konforme Endprodukte über andere Edukte oder Zwischenprodukte bilden. Diesbezüglich können alternative Enzyme, die notwendig sind, damit etwaige Stoffwechselwege unter physiologischen Bedingungen ablaufen, manuell durch den Benutzer ausgemacht werden. Dabei kann unterschieden werden zwischen Enzymklassen, die entweder Reaktionen alternativer Ketten katalysieren oder die, die gleiche Funktionen bei den gleichen Reaktionen übernehmen. Oftmals werden hier Enzyme breiter Substratspezifität aufgeführt. Dies bedeutet, dass diese Enzyme mehr als ein spezifisches Substrat in ihrem aktiven Zentrum binden und umsetzen können, so dass diverse Reaktionen katalysiert werden. [2]

1.10 Die „Modelliererwelt“ verbessern? – die Leitthesen.

Die Problematik der Erstellung metabolischer Modelle in Bezug auf die Richtigkeit mutmaßlicher Enzymannotationen zu den Genen eines untersuchten Organismus ist bereits in Kapitel 1.3 aufgegriffen. Zwar bietet ein Programm wie der EnzymeDetector eine wichtige Hilfestellung bei der Zuordnung wahrscheinlicher enzymatischer Funktionen, dennoch ist die Lösung des Problems unbeschriebener und nicht nachgewiesener Mechanismen, Enzymklassen und Zusammenhänge im Metabolismus von oftmals bisher nicht

eingehend untersuchten Organismen in der Modelliererwelt der Systembiologie zumeist die Annahme von bekannten und fundamentierten Feststellungen bei Vertretern im Mittelpunkt der Forschung, wie zum Beispiel *Escherichia coli*. (Kapitel 1.4) Es ist jedoch nachgewiesen, dass ökologische Umweltaspekte, wie etwa das Vorhandensein von Sauerstoff, fundamentalen Einfluss auf den Metabolismus von Organismen haben und weiterhin die vorwiegende Verfügbarkeit von einzelnen Metaboliten im Habitat ausreicht, um essentielle Differenzen im Reaktionsnetzwerk verschieden evolvierter Organismen nach sich zu ziehen. [29] Dies widerspricht genau der Annahme von Funktionen und Stoffwechselwegen bei Organismen, die im Laufe der Evolution unterschiedliche Habitate mit verschiedenen Umweltbedingungen und Metadaten zu ihrem Lebensraum machten.

Auf diese Begebenheit ist in der Modelliererwelt der Systembiologie vehement hinzuweisen und die Feststellung durch die nachfolgenden Leitthesen dieser Arbeit zu bekräftigen. Der Grundsatz, die Phylogenie sehr wohl bedacht in die Erstellung metabolischer Netzwerke, jedoch auf der Basis von vergleichbaren Metadaten und Umwelteinflüssen verschiedener Organismen einfließen zu lassen, soll am Beispiel dieser Zusammenhänge von Enzymfunktionen und Stoffwechselwegen mit den Metadaten von Organismen der Archaeen-Domäne und einer analysierten Bakterienfamilie der Rhodobacteraceen gefestigt werden. Dies und der Hinweis auf eventuelle neu zu klassifizierende Enzymfunktionen und Stoffwechselwege, die aus diesem Kontext hervorgehen, sollen die Modellierung von metabolischen Netzwerken untersuchter Organismen weiter verbessern.

Die Leitthesen

1. Die Annotation von Enzymen zu einem untersuchten Organismus ist ohne einen sequenzhomologen oder experimentellen Hinweis nur bedingt durch den Vergleich mit nahe verwandten Organismen möglich.
2. Die Untersuchung phylogenetischer Kladen zeigt auf Enzym-, bzw. Stoffwechselwegebene Unterschiede zum allgemein anerkannten 16S rRNA Verfahren.
3. Enzyme und Stoffwechselwege lassen sich mittels Metadaten gruppieren und typische Auffälligkeiten feststellen.
4. Die Untersuchung phylogenetischer Kladen auf Enzym-, bzw. Stoffwechselwegebene kann durch die Zusammenstellung von wiederkehrenden Stoffwechselweglücken und gehäuft vorkommenden Enzymen ohne bisherige Zuordnung neue Enzymvorschläge und neue Stoffwechselwegansätze hervorbringen.
5. Die Archaeen brachten im Laufe der Evolution eigens entstandene Enzyme hervor, die sich nicht zwangsweise in der Funktion zu den Bacteria oder auch den Eukaryota unterscheiden, sich jedoch sequenzhomolog deutlich distanzieren.

1.11 Die Zielsetzungen der Arbeit

Die im vorherigen Kapitel aufgestellten Leitthesen sind in der vorliegenden Arbeit zu prüfen und anhand ausgewählter Verfahren und Beispiele innerhalb der zu analysierenden Organismengruppen zu bestärken (Kapitel 3.3 bis Kapitel-3.11). Ausgangspunkt für sowohl die Erstellung metabolischer Modelle, als auch der hier zu interpretierenden Charakterisierung einer vollständigen bakteriellen Klade und der Archaeen-Domäne in Bezug auf zu recherchierende Metadaten ist in jedem Fall die fundamentierte Annotation von Enzymklassen zu den jeweiligen Genen aller in Betracht gezogenen Organismen. Das entwickelte Programm namens EnzymeDetector soll optimiert, erweitert und korrigiert werden (Kapitel 3.1). Mit dieser Applikation wird ein umfassendes und rationales Bewertungsschema für die Informationen aller in der Bioinformatik etablierten Annotationsdatenbanken angestrebt. Durch den Vergleich der in den verwendeten Datenbanken zu Genen annotierten EC-Nummern und folglich der katalysierten Reaktionen können auch auf Sequenzebene nicht direkt miteinander verwandte Enzyme zusammen gruppiert und in Zusammenhang mit den vorherrschenden Umweltbedingungen in den Lebensräumen gestellt werden. Dies ist im Sinne der Arbeit und ermöglicht ebenso wie die zum Ziel gesetzte Annotation aller sequenzierten Bakterien- und Archaeengenome, ganze phylogenetische Domänen und Familien im Kontext zu analysieren. Essentiell für die Zuordnung von Stoffwechselwegen auf Grundlage von annotierten Enzymen und darauf basierenden Interpretationen ist die Erkennung und Unterscheidung von enzymatischen Schlüsselfunktionen zu alternativen Reaktionswegen, die konforme Substrate und Endprodukte verstoffwechseln. Dementsprechend soll eine Datengrundlage geschaffen und die Problematik in-silico umgesetzt werden (Kapitel 3.2).

2 Material und Methoden

2.1 Die EC-Nummern

Die Basis der in dieser Arbeit angewendeten Verfahren stellt die Zuordnung von Enzymfunktionen zu den Genen der jeweiligen Organismengenome dar. Diesbezüglich wird mit Hilfe der EC-Nummern eine Klassifizierung übernommen, um die Informationen datenbank- und organismusübergreifend einzuordnen.

Durch den Vergleich der in den verwendeten Datenbanken zu Genen annotierten EC-Nummern und folglich nur der katalysierten Reaktionen, können auch auf Sequenzebene nicht direkt miteinander verwandte Enzyme zusammen gruppiert werden. Dies ist im Sinne der Arbeit und ermöglicht die eindeutige Gegenüberstellung der Informationen entweder verschiedener Datenbanken zu einem Organismus oder von Organismen untereinander. Die Problematik der Mehrfachbenennung von Enzymen, die die gleiche Funktion aufweisen, und deren Abgleich wird somit ebenfalls vermieden. In dieser Arbeit werden weiterhin nur eindeutig identifizierte Reaktionsmechanismen, das heißt vollständig deklarierte EC-Nummern, aufgenommen, um Fehlinterpretationen vorzubeugen. Vorläufige Laufnummern für bereits klassifizierte Enzymfunktionen der Datenbanken UniProt und BRENDA werden ebenfalls mit in die Auswertung aufgenommen, ohne dass sie schon bei der IUBMB registriert sind (Kapitel 1.5).

2.2 Weitere bioinformatische Programme

Die Ermittlung der Sequenzähnlichkeit einer Suchsequenz eines Gens oder Proteins zu den Sequenzen einer zuvor zusammengestellten Sequenzdatenbank erfolgt mittels des heuristischen Algorithmus BLAST des NCBI [93] (Kapitel 1.9) (Version 2.2.31+). Die Ergebnisse der so erzeugten lokalen Alignments ermöglichen über den ermittelten Erwartungswert („*Expected (E)-Value*“), die Identität und die Positionsbestimmung deckungsgleicher Module der Vergleichssequenzen ein genaues Ähnlichkeitsmaß selbiger. Die benötigten BLAST-Sequenzdatenbanken zu SwissProt und UniProt [32] (Stand 12/2015) werden durch das Unterprogramm *makeblastdb* erzeugt. Der Proteinsequenzvergleich der einzelnen Gene wird mittels *blastp* erstellt.

Eine weitere Möglichkeit, homologe Sequenzen ausfindig zu machen und infolgedessen gleiche Funktionen zuzuordnen, bietet die Verwendung positionsspezifischer Profile, den sogenannten „*profile hidden Markov models*“ (*profile-HMM*) [94] (Kapitel 1.9, Pfam). Diese Heuristik vergleicht Sequenzen auf Grund eines Statistikmodells, wie wahrscheinlich bestimmte Aminosäuren an bestimmten Positionen vorkommen müssen, bzw. für wie wahrscheinlich eine Substitution, Insertion oder Deletion in diesem Bereich gehalten werden kann. Infolgedessen kann auch so ein Ähnlichkeitsmaß von Vergleichssequenzen mit Hilfe eines Erwartungswertes und der positionsspezifischen Abdeckung, sowie der Identität erhalten werden. In der vorliegenden Arbeit wurde (HMMER Version 3.1b2), genauer *hmmsearch*, verwendet, um Proteinsequenzen gegen die profile-HMM Datenbanken einzelner Proteinfamilien zu vergleichen.

2.3 Die phylogenetischen Applikationen

Einige wichtige grundlegende Verfahrensweisen der Phylogenie finden in dieser Arbeit Anwendung, um Zusammenhänge des Metabolismus der Organismen mit der Evolutionsgeschichte zu identifizieren oder Differenzen zur Einordnung der systematischen Biologie zu erkennen.

Die Darstellung der Verwandtschaftsverhältnisse untersuchter Organismen erfolgt in sogenannten Kladogrammen. Hier entspricht jede Verzweigung mindestens einem hinzugekommenen evolutionären Merkmal. Diese können phänotypisch und morphologisch sein, oder beispielsweise auf Differenzen der Gensequenzen basieren. Infolge können randständig die Taxa als Blätter des so erzeugten Baums, das bedeutet die einzelnen Organismen, abgelesen werden. Dabei weist jedes Taxon zu jedem anderen Organismus eine entsprechende Distanz durch Merkmalsdifferenzen auf. [14]

Um eine Organismengruppe in diesen Verhältnissen darstellen zu können, werden oft definitiv außenständige Organismen, die eine größere Distanz zu allen weiteren Taxa aufweisen, der Untersuchung hinzugefügt. Infolgedessen besteht laut der Evolutionsgeschichte ein gemeinsamer Vorfahre der Organismen von Interesse und der ergänzten sogenannten *Outgroup*. [14] Um dies zu verbildlichen, wird dem Baum eine Wurzel eingefügt. Ein bioinformatisches Verfahren, um dies zu erkennen, wird als *Midpoint-Rooting* bezeichnet. Hier entspricht die Hälfte der größten im Baum zu findenden Distanz der Position der Wurzel. Diese Position sollte per Definition folglich zwischen der *Outgroup* und der betrachteten Organismusgruppe liegen. Diese Tatsache gilt als Qualitätsmerkmal der aufgestellten Verwandtschaftsverhältnisse. [95]

Um die Ähnlichkeit von Organismen auf der Basis von Sequenzen, wie zum Beispiel der 16sRNA, oder auf der Grundlage von Eigenschaften, wie dem Vorhandensein von bestimmten Enzymen oder metabolischen Stoffwechselwegen (Kapitel 2.9, Kapitel 3.10, Kapitel 3.11), zu quantifizieren und in phylogenetischen Bäumen grafisch darzustellen, wird das Online-Framework Mobyle @Pasteur [96] verwendet (Abbildung 1, Kapitel 2.9). Die Sequenzen der 16S rRNA Untereinheiten wurden für die Rhodobacteraceen vom NCBI [97] einzeln extrahiert. Bei den Archaeen dient das „NCBI RefSeq Targeted Loci Project“ [98] als direkte Downloadquelle der benötigten RNA-Sequenzen. Mit Hilfe des Multiplen Alignment-Programms MUSCLE (Version 3.8.31) [99] wird die Grundlage für die Berechnung einer Distanzmatrix mittels *dnadist* (PHYLIP Version 3.67) [100] erstellt. Mit einem weiteren Programm der PHYLIP Bibliothek namens *neighbor* werden darauf basierend phylogenetische Bäume nach dem Neighbor-Joining Algorithmus [101] generiert. Als Parameter werden 100 Bootstrap Replikate erzeugt, die Inputreihenfolge (Jumble) randomisiert und die entsprechende *Outgroup* angegeben. Die Replikation des Verfahrens mit unterschiedlicher Startreihenfolge der Organismen, ist ein weiteres Qualitätskriterium phylogenetischer Bäume. Verwandtschaften und Verzweigungen, die sich in allen Replikaten gleich darstellen, scheinen demzufolge entsprechend gefestigt. [95]

Das Neighbor-Joining Verfahren nimmt das „Minimum Evolution“-Kriterium als Grundlage, um ausgehend von einem zunächst „sternenförmigen“ phylogenetischen Baum immer die beiden ähnlichsten Taxa, bzw. Cluster zu

einem Ast zusammenzufassen. Dabei ist eine konstante Evolutionsgeschwindigkeit nicht Voraussetzung. [101]

Wie eingangs erwähnt können Sequenzunterschiede dazu verwendet werden, um Distanzen zwischen den jeweiligen Genomen zu analysieren. Diese Differenzen können allerdings auch durch Merkmale, wie die codierten Enzyme und die dem Metabolismus der Organismen zugeordneten Stoffwechselwege hervorgebracht und untersucht werden. Hierzu wird ein sogenannter Binärvektor erzeugt, der an entsprechender Position, die für ein Enzym, bzw. einen Stoffwechselweg steht, eine 1 für vorhanden und eine 0 für nicht vorhanden vergibt. Aus diesen Vektoren wiederum kann durch einfaches zählen der Unterschiede die Distanz berechnet werden (Hamming-Distanz [102]). Liegen diese Daten nicht für eine systematisch außenstehende Randgruppe von Organismen vor, kann ein als Positivvektor bezeichneter „Organismus“ simuliert werden. Ihm werden alle Merkmale, die die untersuchten Taxa unterscheiden, zugesprochen (Binärvektor besteht nur aus Einsen), sodass er die größte Distanz zu allen Organismen aufweist.

Zur Darstellung der phylogenetischen Bäume und visuellen Bearbeitung dient PhyloWidget [103].

2.4 Die Sequenz-, Annotations-, Stoffwechselweg- und Metadaten

Der EnzymeDetector Algorithmus extrahiert aus diversen Quelldatenbanken Enzymannotationen aller sequenzierten Genome der Archaeen und Bakterien. Um die Zuordnung zu den jeweiligen Genloci zu gewährleisten, wird bei einigen Datenbanken eine explizite Datendatei angeboten, um Sequenzbezeichnungen auf den entsprechenden Genlocus abzubilden. Die heruntergeladenen Datendateien werden per Python-Skript verarbeitet und die relevanten Informationen in einer MySQL-Datenbank gespeichert (Anhang 5).

PEDANT (Kapitel 1.9) stellt seit Beginn der Weiterentwicklung des EnzymeDetector 08/2012 keine aktualisierten Daten zur Verfügung. Somit werden neu sequenzierte Organismen ebenso wenig beachtet, wie neue Proteinfamilien und Kategorisierungen zur Funktionsvorhersage. Aus diesem Grund wird diese Datenbank aus der vorliegenden EnzymeDetector Analyse ausgeschlossen.

NCBI

Es werden für die aktuelle Version des EnzymeDetector (Kapitel 1.6) mehr als 3000 finale bakterielle und archaeale Genome und infolgedessen insgesamt mehr als 7000 Assemblies zu Chromosomen und Plasmiden mit den entsprechenden Enzymannotationen des NCBI (Stand 12/2015, Kapitel 1.9) extrahiert. Dazu erfolgt der FTP Download der Datendateien der jeweiligen Genome. Die Informationen des Genlocus der Sequenzen werden in dieser Datei mit aufgeführt und vom erstellten Skript gespeichert.

Weiterhin wird die Sammlung aller nicht redundanten Gensequenzen, die „*non-redundant*“ (nr)-Datenbank, genutzt, um die Sequenzinformationen zu ca. 230 Millionen Genen zuzuordnen (Stand 12/2015, Kapitel 1.9).

KEGG

Die KEGG Datenbank führt Enzymannotationen zu 3736 Organismen. Darunter werden 1,3 Millionen Annotationen aus dem Genomprojekt, sowie 3,3 Millionen Enzymzuweisungen zu Genen aus der KEGG-Orthology gewonnen (Stand 12/2015, Kapitel 1.9). Die Zuordnung der Genloci wird in einer separaten CSV-Datei angeboten und ausgewertet.

PATRIC

Etwa 20 Millionen Enzymannotationen zu 5383 Bakterien und Archaeen zugeordnet liefert die per FTP heruntergeladene Datendatei (Stand 12/2015, Kapitel 1.9). Der Bezug zu den vom Sequenzierungsprojekt vorgegebenen Genloci wird direkt mitgeführt und verarbeitet.

UniProt/SwissProt

Die Sequenzen der SwissProt-Datenbank können etwa 4300 Enzymklassen zugeordnet werden (Stand 12/2015, Kapitel 1.9). Diese Annotationen dienen einerseits selbst als Datenquelle für den EnzymeDetector und andererseits der Funktionsbestimmung homologer Sequenzen durch den BLAST-Algorithmus. Die Abbildung der individuellen sequenzeindeutigen Bezeichner (*UniProtID*) auf die Genloci erfolgt durch die Verarbeitung einer separaten Datendatei.

Pfam

Mit Hilfe der Pfam-Datenbank können unbekannte Proteinsequenzen in ca. 16000 verschiedene Proteinfamilien auf Grund des Vorkommens bestimmter Domänen und deren Architektur über die Sequenz verteilt eingeordnet werden (Stand 11/2015, Kapitel 1.9). Zu allen Proteinfamilien verknüpft die Datenbank ein multiples Sequenzalignment und ein HMM. Ordnet man diesen Profilen, die demzufolge für bestimmte Proteine stehen, nun Enzymfunktionen, die aus sicherer Annotation bekannt sind, zu, erhält man eine Möglichkeit, unbekannte Sequenzen zu annotieren [104].

BREPS

Die offline verfügbare MySQL-Datenbank auf dem Stand 03/2015 (Kapitel 1.9) führt über 15600 Sequenzmuster zu mehr als 4000 verschiedenen Enzymen, bzw. Enzymkombinationen für die ein *Pattern* steht. Per Python-Skript (Anhang 5) werden die Sequenzen der Organismen mit diesen Mustern abgeglichen und die Übereinstimmung entsprechend bewertet (Kapitel 1.6 und Kapitel 1.9).

BRENDA

Die BRENDA ermöglicht die Zuordnung von etwas mehr als 6000 Enzymklassen zu knapp 20000 verschieden deklarierten Organismen und Organismen-Stämmen aller Domänen. Diese Informationen werden aus der innerhalb der Arbeitsgruppe verfügbaren MySQL-Datenbank vom Stand 01/2016 (Kapitel 1.9) extrahiert. Die geführten Sequenzbezeichner entsprechen denen der UniProt-Datenbank. Somit können diese mit den dort erhaltenen Mapping-Informationen auf Genloci abgebildet werden.

Wichtige Informationen und Literaturverweise zu Enzymdaten und -funktionen werden zur Interpretation mit Hilfe des Webservice nachgeschlagen. Zusammenhänge der Enzymklassen mit entsprechenden Stoffwechselwegen werden ebenfalls mit den seit dem Release 01/2016 angebotenen *Pathway*-Karten nachvollzogen und mit anderen Quellen online verglichen.

AMENDA

Das automatisierte Textmining verknüpft 4643 Enzyme mit mehr als 28000 verschiedenen Organismen- und Stammbezeichnungen zu über 387000 Annotationen. Dabei wird die MySQL-Datenbank auf dem Stand 01/2016 (Kapitel 1.9) abgefragt und die Kategorisierung der Qualität zwischengespeichert und zur Konfidenzbewertung herangezogen.

MetaCyc

Von MetaCyc werden 2260 Pathways von 2600 verschiedenen Organismen aus den Domänen der Bakterien, Archaeen und Eukaryoten sowohl online dargestellt, als auch zum Herunterladen als Datenfiles zur Verfügung gestellt. (Stand 01/2016) Die jeweiligen Verknüpfungen von Metaboliten durch Reaktionen und deren katalysierenden Enzymen können durch das angegebene *reaction-layout* nachvollzogen werden. Dieses wird ebenfalls auf dem Stand 01/2016 heruntergeladen und in eine MySQL-Datenbank überführt. [105] Somit wird die Grundlage geschaffen, automatisiert ganze Stoffwechselwege abzubilden und so die obligatorischen Reaktionsschritte inklusive der zugehörigen Enzyme von alternativen Seitenzweigen oder Aufspaltungen im Reaktionsnetz zu erkennen. Die zugehörige Angabe des sogenannten „Pathwaylayouts“ in Form von Reaktionsketten der Substrate, Zwischen- und Endprodukte, ermöglicht eine automatisierte Unterscheidung beteiligter enzymatischer Schlüsselfunktionen, für die dieser Stoffwechselweg keine Alternativen vorsieht, von Enzymen, die über unterschiedliche Reaktionen dennoch zu konformen Produkten führen. Das hierfür erstellte Programm (Anhang 5) ist essentiell und von hoher Bedeutung für die Auswertung auf Ebene der Stoffwechselwege. Infolgedessen werden explizit nur Pathways den Organismen basierend auf ihrem Enzympool zugeordnet, die ihr Metabolismus funktionell bedienen kann.

Hierbei wird, sobald ein Zwischen- oder Endprodukt aus zwei voneinander verschiedenen Reaktionen resultiert, per Rekursion nachvollzogen, ob die Rückwege wiederum zu einem gemeinsamen Edukt führen. Ist dies der Fall, liegen alternative Wege vor. Andernfalls werden für das Endprodukt Verbindungen als Teilbausteine geliefert, sodass diese obligatorisch durch die entsprechenden Enzyme katalysiert ablaufen müssen. Zyklische Stoffkreisläufe werden durch das Zwischenspeichern aller bereits abgearbeiteten Metabolite erkannt und demgemäß in der Datenbank abgebildet. Lineare Reaktionen von direkt verfügbaren Substraten zu einem Produkt werden durch die entsprechenden obligatorischen Enzyme katalysiert. Von MetaCyc direkt angegebene alternative Enzyme zur Katalyse gleicher Reaktionen werden als Alternativen verarbeitet. Zu jedem durch das *reaction-layout* nachvollzogenen Stoffwechselweg werden in einer MySQL-Datenbank (Anhang 5) die erhaltenen obligatorischen Enzymvektoren in einer Klammerschreibweise gespeichert

([Enzym 1] - [Enzym 2] – [...]). Dabei entspricht eine eckige Klammerung immer einem der Reaktionsschritte, die in diesem Fall allerdings nicht geordnet vorliegen müssen. Bei der Auswertung der Enzym pools von Organismen muss jede dieser Enzymklassen annotiert sein.

Anschließend werden die alternativen Vektoren nacheinander in der MySQL-Datenbank aufgeführt. So steht jedem Enzymvektor, der hier einen zusammenhängenden Teilstück des Stoffwechselweges beschreibt, ein Alternativvektor von Reaktionsketten, die zu gleichen Produkten führen, gegenüber ([Enzym 1A] - [Enzym 2A] – [...] --- [Enzym 1B] - [Enzym 2B] – [...]). Von jeder dieser Gegenüberstellungen müssen bei der Auswertung mindestens einer der beiden Vektoren und seine Enzyme dem Organismus zur Verfügung stehen.

Direkte Alternativen für ein und dieselbe Reaktion werden mit der Klammerschreibweise ebenfalls dargestellt und der Auswertung genügt die Annotation eines der beiden Funktionen ([Enzym 1.1 - Enzym 1.2] – [...]).

Den Verfahren dieser Arbeit (Kapitel 2.9) steht somit eine Beschreibung von Stoffwechselwegen hinsichtlich der zwingend benötigten enzymatischen Funktionen und alternativer Enzyme zur Verfügung, um definitiv auf der Basis der erstellten Annotation der Organismen über die Zuordnung der Pathways zu entscheiden (Anhang 5).

Metadaten und mikrobielle Eigenschaften

In den Analysen dieser Arbeit wird geprüft, ob ein Zusammenhang zwischen dem Vorkommen von Enzymfunktionen und den zu sammelnden Metadaten der untersuchten Organismengruppen besteht. (Kapitel 2.9) Dazu ist die Festlegung der zu unterscheidenden mikrobiellen Eigenschaften ebenso essentiell wie Literaturquellen zu diesen. In der Gruppe der Archaeen wird eine reine Literaturrecherche [106] im Rahmen des *SulfoSYS* Projektes mit der TU Braunschweig als Kooperationspartner durchgeführt und es können die jeweiligen Merkmale aus der Tabelle 2 typisiert werden. Unterschieden wird beispielsweise nach dem pH-Wert Optimum der Vertreter. Hierfür werden Bereiche zugeordnet, die in der späteren Auswertung für acidophile (pH0-5) oder alkaliphile (pH9-14) Archaeen stehen. Ebenso verhält es sich mit dem recherchierten Temperaturoptimum. Hier werden psychrophile Vertreter (0-41°C) den thermophilen (90-100°C) gegenübergestellt. Der Sauerstoffmetabolismus wird Organismen zugeteilt, die zwingend Sauerstoff benötigen, fakultativ auch einen lebensfähigen Metabolismus ohne oder gänzlich unter anoxischen Bedingungen nachweisen. Gesondert gekennzeichnet werden Archaeen, denen in der Literatur die Fähigkeit zu Energiegewinnung unter Methanbildung zugesprochen wird. Insofern die Informationen verfügbar sind, wird die Art des Metabolismus nach chemo-, lito-, organo-, auto- und heterotroph differenziert. Die Halotoleranz der Archaeen ist charakterisiert durch die Natriumchloridkonzentration, in der sie überlebensfähig und auch anzutreffen sind. Demgemäß wird unterschieden in gering (1-6%NaCl-Konzentration), medium (6-15%NaCl-Konzentration) und extrem halotolerante Vertreter (>15%NaCl-Konzentration). Die verschiedenen Habitate, die in der Literatur beschrieben sind, gliedern sich auf in terrestrische Archaeen, die beispielsweise im Erdboden oder auf der Erdoberfläche assoziiert mit anderen Organismen leben und aquatische Vertreter, die wässrige Habitate beleben. Eine weitere

2 Material und Methoden

Gruppe an Populationen sind im Schmutz-, Schlick- und Abwasser zu finden. Hier herrschen oft anaerobe Bedingungen, die zudem noch mit toxischen Substanzen aufwarten können. Besonders thermophile Archaeen leben in sogenannten heißen Quellen und Schwefelquellen oder auch in vulkanischen Gebieten. All diesen Habitaten ist die hohe Umgebungstemperatur gemeinsam. Dabei können die heißen Wasserquellen durch unterirdische vulkanische Aktivitäten erhitztes Wasser an die Erdoberfläche befördern. Ebenso können sie am Meeresgrund auftreten und sogenannte *White Smoker* bilden, während Temperaturen bis um die 80°C erreicht werden. Wesentlich höhere Temperaturen erreichen die Schwefelquellen (*Black Smoker*, über 100°C). In ihrem austretenden heißen Wasser sind vor allem Sulfide und andere Salze von Eisen, Mangan, Kupfer und Zink gelöst. Sie treten ebenfalls auch an der Erdoberfläche auf und sind dort für den Menschen vornehmlich an ihrem typischen fauligen Geruch zu erkennen und von besonderer Gefahr [14]

Tabelle 2: Typisierte Metadaten und ihre Kategorien in den Projektorganismengruppen. Die Recherche der verfügbaren Daten führt bei den Rhodobacteraceen zu den Metadaten der Unterscheidung des marinen und nicht marinen Lebensraumes, der Isolationsart und dem Lebensstil. Bei den Archaeen können neben den Metadaten pH- und Temperatur-Optimum der Organismen noch Differenzierungen bezüglich des Sauerstoff- und Methanmetabolismus gemacht werden. Weiterhin dienen der generelle Metabolismustyp, das Habitat und die Salztoleranz der Vertreter als Unterscheidungskriterium. Die Salztoleranz wird gegliedert in gering (1-6% NaCl-Umgebung), medium (6-15% NaCl-Umgebung) und extrem (>15% NaCl-Umgebung) halophil.

Projektorganismengruppe: Rhodobacteraceae	
Metadaten	Kategorien
mariner Lebensraum	marin lebend, nicht marin lebend
Isolationsart	freilebend, nicht freilebend (Laborstamm)
Lifestyle	Biofilm, Frischwasser, organismengebunden, Meereis, Sediment, Erdreich, Wasser
Projektorganismengruppe: Archaeen	
Metadaten	Kategorien
pH-Optimum	von, bis
Temperatur-Optimum	von, bis
Sauerstoffmetabolismus	aerob, anaerob, fakultativ
Habitat	terrestrisch, aquatisch, heiße Quellen, vulkanisch, Schwefelquellen, Schmutzwasser
Halotoleranz	gering (1-6%NaCl), medium (6-15%NaCl), extrem halophil (>15%NaCl)
Metabolismus	chemotroph, litotroph, organotroph, autotroph, heterotroph
Methanmetabolismus	methanogen

In der Projektorganismengruppe der Rhodobacteraceen werden Umweltdaten und weitere Angaben der einreichenden Wissenschaftler in Bezug auf die Isolation der Organismen an das NCBI BioProject [97] ausgewertet. So können hier die entsprechenden Metadaten aus der Tabelle 2 zur Charakterisierung herangezogen werden.

Die marinen Vertreter stellen den Großteil der Bakterienfamilie dar und sind in allen Bereichen der Weltmeere, vornehmlich im Oberflächenwasser, zu finden. [61,62] Unter den nicht marinen Rhodobacteraceen werden sowohl bodenlebende Organismen, als auch Vertreter aus Ab- und Brackwasser oder im Schlamm und Sediment von Süßgewässern lebende Bakterien zusammengefasst. Weiterhin soll untersucht werden, ob eventuelle Differenzen

durch den Verlust oder der Evolution von Genen zwischen freilebend isolierten Organismen und Laborstämmen auszumachen sind. Außerdem erfolgt eine Aufgliederung nach dem jeweiligen Lebensstil der Organismen. Demgemäß wird die Kategorie der organismusgebundenen Vertreter für die zumeist mit Algen in Symbiose lebenden Rhodobacteraceen gruppiert (Kapitel 1.8). Demgegenüber stehen Bakterien der Lebensräume im Sediment, im Erdreich und im Wasser. Obwohl die Biofilmbildung in der Literatur als familientypisch beschrieben ist, können nur sehr wenige Vertreter (4) aus den Isolationsprojekten deklariert werden. Dies kann daran liegen, dass diese Organismen aus praktischen Gründen vornehmlich nicht im Zustand des Biofilms, sondern freilebend und nicht aggregiert isoliert werden. Ebenso wenige Organismenbeispiele werden für das Frischwasser (7) und das Meereis (2) recherchiert. Diese nur wenig belegte Habitatgruppen werden auf Grund der mangelnden statistischen Aussagekraft von der Bewertung ausgeschlossen.

2.5 Das High Performance Computing (HPC)-Cluster

Um die Verarbeitung der Genome der angestrebten Organismenanzahl in praktikabler Rechenzeit zu ermöglichen, wird der EnzymeDetector-Algorithmus komplett auf ein HPC-Cluster transferiert und optimiert. Dabei werden sowohl die vorberechneten Annotationsdatenbanken der Quellen auf einem gemeinsamen MySQL-Server erzeugt, als auch die Ergebnistabellen zusammengeführt und bereinigt/komprimiert für den Webserver zur Verfügung gestellt. Der Dateiaustausch im Clusternetzwerk erfolgt über ein gemeinsames Home-Storage im RAID50-Verbund (4,8TB).

Als Betriebssystem der Clusterumgebung wird Debian GNU/Linux 7.9 (wheezy) [107] in der 64Bit Architektur gewählt. Die Topologie ist gekennzeichnet durch einen Masterknoten, 18 Rechenknoten, einem 1Gbit Netzwerk-Master- und einem Management-Switch. (Spezifikationen Tabelle 3)

Das Netzwerkmanagement inklusive IP-Adressvergabe, die Anbindung an das Institutsnetzwerk und das Routing zum Universitätsnetzwerk übernimmt der Masterknoten als DHCP Server. Die Installation des Betriebssystems und die Konfiguration der OS-, User- und Softwareumgebung der komplett autonomen Rechenknoten wird mittels FAI [108] gesteuert. Eine Jobverteilung im Sinne von Multitaskverfahren/-Anwendungen und das Management der Rechenressourcen ist mittels der Stapelverarbeitungssoftware „Son of Grid Engine“ [109] realisiert.

Tabelle 3: Spezifikationen der HPC-Cluster Topologie. Die Netzwerkkommunikation der Knoten untereinander und zum Masterknoten wird von einem Master-Switch 1Gbit übernommen. Auf Grund der simultanen und parallelen Netzwerkanforderungen des Masterknotens wird dieser modular mit 10Gbit angebunden. Ein Management-Switch dient zur externen Erreichbarkeit und zur allgemeinen Serversteuerung und Monitoring. Das HPC-Cluster umfasst zehn Rechenknoten Typ I und acht mit den Spezifikationen des zweiten Typs.

Einheit	Spezifikationen
Master-Switch	HP Procurve, 1Gbit, 10Gbit-Modul zum Masterknoten
Management-Switch	DLink-dgs-3100 (Layer 2), Management Knoten über IPMI Schnittstellen
Masterknoten	6x4GB DDR3 RAM 2xIntel® Xeon® QuadCore HT E5620 [2,4GHz]
Rechenknoten Typ I	6x2GB DDR3 RAM 2xIntel® Xeon® QuadCore HT E5620 [2,4GHz]
Rechenknoten Typ II	8x2GB DDR2 RAM 2xDual-Core AMD Opteron™ Processor 2212 [2GHz]

2.6 Der EnzymeDetector

Das entwickelte Programm namens EnzymeDetector [23] (Kapitel 1.6) wird in der vorliegenden Arbeit verbessert, erweitert und korrigiert. Mit dieser Applikation wird ein umfassendes und rationales Bewertungsschema für die Informationen aller in der Bioinformatik etablierten Annotationsdatenbanken angestrebt. Zusätzlich werden diese durch eine eigens durchgeführte BLAST [93]-Annotation unterstützt oder gegebenenfalls widerlegt.

In der aktuell entwickelten Version des EnzymeDetector und der vorliegenden Arbeit wird die BLAST-Version 2.2.31+ verwendet. Das Teilprogramm der BLAST-Bibliothek *makeblastdb* wird mit Standardparametern dazu genutzt, eine Proteindatenbank von SwissProt [32] (Stand 12/2015) zu erzeugen. Die Analyse durch den Vergleich zur vollständigen UniProt-Datenbank wird durch den Sequenzvergleich zu der Untermenge geprüfter und experimentell belegter Annotationen der SwissProt-Datenbank ersetzt. Dadurch werden häufige Fehlinterpretationen der teilweise rein automatisiert erstellten Zuordnungen von Enzymklassen der TrEMBL-Datenbank ausgeschlossen und ein exakteres Ergebnis erwartet.

Für die erweiterte Analyse der BLAST-Auswertung und vor allem zur Differenzierung von multifunktionellen auf einem Gen codierten Enzymen, werden die BLAST-Ergebniswerte der Identität als Maß für die absolute Übereinstimmung der gegenübergestellten Aminosäuresequenzen und die Positionsangaben übereinstimmender Sequenzabschnitte der Such- und Treffersequenz übernommen und in den Ergebnistabellen weitergeführt. Damit steht eine Möglichkeit zur Verfügung, mehrere Enzymfunktionen auf einem Gen, die nacheinander codiert sind, zu erkennen. Ebenso erfolgt ein Vermerk, sobald als sehr ähnlich eingestufte Sequenzen in SwissProt schon mit mehreren Funktionen erkannt und beschrieben sind.

Die BLAST-Ergebniswerte der Suchsequenz des zu annotierenden Gens werden pro zuzuordnender Enzymklasse von der längsten Übereinstimmung der Datenbanksequenz mit der höchsten Identität und dem besten *E-Value* übernommen. Der Schwellenwert, dass mindestens 5 voneinander verschiedene Treffersequenzen für die jeweilige Annotation ausfindig gemacht werden, bleibt bestehen. Sollte dieser allerdings nicht mindestens einmal erreicht werden, erfolgt dennoch eine Übernahme der erhaltenen Ergebnisse mit entsprechender Qualitätsabwertung, auch wenn weniger als 5 Sequenzbeispiele in SwissProt bekannt sind. Damit werden auch Enzymannotationen zugelassen, die bisher nur selten und bei wenigen Spezies beschrieben sind, während der Hinweis auf eine gesonderte Interpretation durch niedrigere Konfidenzen gewahrt bleibt.

Weiterhin wird in den Ergebnistabellen aufgeführt, inwiefern eine enzymatische Funktion des Genlocus zu erwarten ist. Dazu erfolgt ein Vermerk, sollte der BLAST vorrangig Kongruenzen nach den eben genannten Kriterien bezüglich der Länge und der Identität der Übereinstimmung mit geringerem Erwartungswert bei nicht enzymatisch deklarierten Proteinen der SwissProt-Datenbank ausfindig machen.

Neben diesen sequenzbasierten Annotationsverfahren bindet der EnzymeDetector weiterhin eine Sequenzmustersuche mittels BrEPS [36] (Stand 06/2015) zur Bewertung der Enzymzuordnungen ein. Die verschiedenen Einstufungen der Konfidenz von 1-10 abhängig von der zugeordneten Korrektheit

und Spezifität der Sequenzmuster zur Erkennung von Enzymklassen dieser Datenbank bleiben bestehen (Kapitel 1.6).

Weitere Datenbanken, Quellen und Ansätze für Enzym-Annotationen werden gefunden und eingebunden. Dies soll den Vorteil gegenüber einzelnen Datenbanken bezüglich einer optimierten Annotation von Genen weiter ausbauen. Dabei sollen jedoch auch weiterführende Informationen bereits inkludierter Datenbanken einfließen. Dementsprechend werden die Bewertung und Qualitätseinstufung konkurrenzbasierter Zusammenhänge von Enzymklassen und Organismen des automatisierten Textminings AMENDA (Kapitel 1.9) ebenso beachtet, wie die Differenzierungen und Informationen der BRENDA (Kapitel 1.9) zu Annotationen der Gene von Organismusstämmen auf Sequenzebene oder ohne eine Verknüpfung zu den Genloci.

Dem Wunsch der Anwender und der Anforderung der folgenden Arbeit nach der Möglichkeit der automatisierten und möglichst optimierten Annotation kompletter Genome und in Folge auch ganzer phylogenetischer Kladen (Kapitel 2.9: Vorgehen nach Verfahren II) wird mit all dem ebenso Folge getragen, wie mit einer grundlegend überarbeiteten Version des frei zugänglichen Webinterfaces. Dabei stehen der schnelle Überblick, die komfortable Filterung der Ergebnisse und der Zugang zu einem automatisiert verarbeitbarem Format im Vordergrund.

Die Untersuchung ganzer phylogenetischer Domänen bzw. Gruppen der Bakterien und Archaeen setzt eine umfassende Abdeckung sequenzannotierter Organismen voraus. Dem wiederum gehen eine Verarbeitung von Datenbankinformationen und Eigenannotationen der Quellorganismen in praktikabler Rechenzeit voraus. In diesem Zuge wird der Algorithmus des EnzymeDetector überarbeitet und ein Multitask-Verfahren auf einem „*High Performance Computing*“-Cluster (Kapitel 2.5) installiert und optimiert, um eine vollständige Annotation aller final sequenzierten Genome der Bakterien und Archaeen in geeigneter Rechenzeit für einen halbjährlichen Updatezyklus zu gewährleisten.

Die Zuordnung von Enzymklassen zu entsprechenden Genen aus verschiedenen Datenbankquellen wird modular umgesetzt und auf Grundlage des Genlocus eingeordnet. Dieser Genlocus entspricht einer festgelegten Position des Gens auf dem Genom eines Organismus und ist von Beginn der Genzuweisung nach der Sequenzierung der Spezies definiert. Somit wird jede etwaige Annotation von Enzymfunktionen einer Quelle der Genomposition zugeordnet insofern sie vorliegen. Dieses Verfahren ist ohne Bezug auf organismusunspezifische und rein sequenzeindeutige Zuordnungen wie den variablen *Geneidentifier* des NCBI und somit für die Abbildung von Annotationen verschiedener Quellen auf Gene vorzuziehen.

Der erstellte modulare Aufbau des Algorithmus unterscheidet die Schritte des Extrahierens der Annotationsinformationen der jeweiligen Datenbanken zur Überführung in eigene Datentabellen zur eindeutigen, quellspezifischen Zuordnung von Enzymen zu Genloci von der letztendlichen Analyse der Gene der untersuchten Organismen. Dies ermöglicht einerseits ebenso die einfache Integrierung weiterer Quellen in Zukunft, sowie andererseits die schnell verfügbare Information datenbankspezifischer Annotationen für die auszuwertenden Gene durch eine nicht redundante und indexoptimierte MySQL-Datenbanktabellenstruktur.

Die Annotation, hervorgebracht durch das Sammeln und Vergleichen von Funktionszuweisungen eines Gens, ist direkt abhängig von der Bewertung einzelner Quellen. [23] Das Bewertungsschema des EnzymeDetector wird daher erneut überprüft und gegebenenfalls angepasst.

Das multitask-Verfahren wird einerseits auf Ebene der zu verarbeitenden Organismen und andererseits auf Ebene der zu berechnenden BLAST- und HMM-Jobs (Kapitel 2.2) angewendet, um eine optimale Zeitersparnis zu erreichen.

2.7 Das EnzymeDetector-Webinterface

Die Ergebnisse des EnzymeDetector werden frei zugänglich auf einem weiterentwickelten und den Anforderungen entsprechendem Webinterface unter <http://enzymedetector.tu-bs.de> zur Verfügung gestellt. Zur Verwaltung des Seitenaufbaus, des Inhalts und des grafischen Designs wird das Framework Wordpress 4.3 [110] verwendet. Für die Darstellung der Organismus-, Enzym- und Ergebnistabellen werden JavaScript-Elemente aus der frei zugänglichen Bibliothek namens jQuery [111] verwendet und angepasst. Die serverseitige Verarbeitung von User-Anfragen und die Schnittstelle zum Datenbanksystem werden mittels PHP [112] umgesetzt. Clientseitig werden das Design und die Anordnung des Inhalts durch HTML-Code [113] und **Cascading Style Sheets** [113] vorgenommen.

2.8 Weitere Software

Die Speicherung und Verarbeitung der Ergebnisse und Zwischenergebnisse des Projektes und des EnzymeDetector erfolgt in einem SQL-Datenbanksystem verwaltet durch MySQL [114,115]. Python [116] ist die Programmiersprache aller entwickelten Quellcodes des EnzymeDetector und der Anwendungen zur Auswertung der vorliegenden Arbeit. Als Python-Datenbank-API dient MySQLdb [117]. Die Zusammenfassung und die Aufbereitung von Rohdaten zur Findung von Interpretationsansätzen, zum Beispiel durch Farbskalen, wird mit Microsoft EXCEL 2016 [118] bewerkstelligt.

2.9 Die Auswerteverfahren metadatentypischer Enzyme/Stoffwechselwege

Die vorliegende Arbeit verfolgt zwei Ansätze, um mögliche Enzyme, die in den untersuchten Organismengruppen auffallend typisch für bestimmte Metadaten stehen, zu identifizieren.

Das Vorgehen nach Verfahren I

Ein erstes Verfahren macht sich die Sammlung aller Enzyme, denen die BRENDA-Datenbank (Kapitel 1.9) Sequenzen zugeordnet hat, zu Nutze (Abbildung 1 und Kapitel 3.4). Alle diese werden gegen die UniProt-TrEMBL-Datenbank geblastet (Kapitel 1.9), um potentielle Organismen zusammenzutragen, die mit höchster Wahrscheinlichkeit auch die jeweils betrachtete Enzymfunktion in ihrem Metabolismus vorweisen können. Als Grenzwerte der BLAST-Analyse werden ein *E-Value* geringer als 10^{-3} und ein Identitätswert größer als 45% angesetzt. Es werden nur die erarbeiteten Metadaten der Archaeen in diesem Ansatz ausgewertet. Fallen in die Ergebnismenge große Anteile in eine der in Kapitel 2.4 beschriebenen Metadatenkategorien, kann dies als Hinweis auf einen Zusammenhang mit der jeweils betroffenen Enzymfunktion zu sehen sein. Diese Auswertung der Ergebnisdatentabellen erfolgt mittels Excel und der Hervorhebung durch bedingte Farbformatierungen. Dabei folgt eine Farbskala den Anteilen der jeweils annotierten Enzyme einer Metadatenkategorie. Somit können auffällige Zusammenhänge schnell erkannt werden. (Anhang 3) Die Ergebnisse der zuvor erstellten BLAST-Vergleiche werden aus den in einer MySQL Datenbank zwischengespeicherten Informationen per Python-Skript extrahiert (Anhang 5). Im Anschluss soll die Frage nach Enzymen, die innerhalb der Archaeen evolutionär gesehen höchstwahrscheinlich entstanden sind, beleuchtet werden. Dies setzt keine vollständig neue Enzymfunktion in der Domäne voraus. Ebenso sollen archaeentypische Enzyme erkannt werden, die auf Sequenzebene eine deutliche Distanz zu anderen annotierten Vertretern aus anderen Domänen wie zum Beispiel den Bakterien aufweisen. Dazu wird die Anzahl der erkannten BLAST-Treffer der Domänen Archaeen und Bakterien im Verlauf zum Erwartungswert gegenübergestellt. Der Median des prozentualen Anteils der erkannten Organismen zu der insgesamt bekannten Organismenzahl der entsprechenden Domäne über die *E-Value*-Stufen $<10^{-200}$, 10^{-200} , 10^{-180} , 10^{-160} , 10^{-140} , 10^{-120} , 10^{-100} , 10^{-80} , 10^{-60} , 10^{-40} , 10^{-30} , 10^{-20} , 10^{-10} , 10^{-3} dient als Kriterium, um domäentypische Enzyme zu bewerten. Archaeentypische Enzyme sollen eine Mediandifferenz zu den Bakterien von mindestens 40 aufweisen. Hingegen charakterisiert mindestens eine Differenz von 50 bakterientypische Enzyme. Die Berechnungen und die Markierung entsprechend hoher Differenzen durch bedingte Farbformatierungen erfolgen in Excel. (Anhang 4)

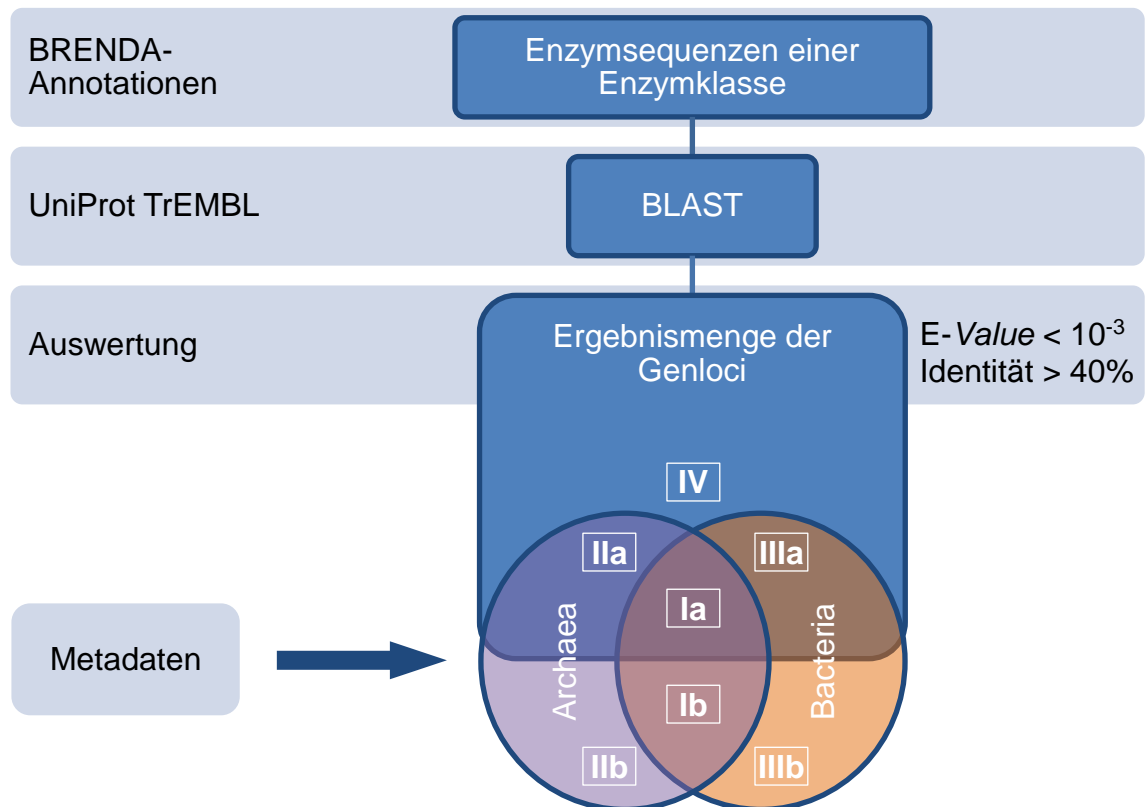


Abbildung 1: Schema Verfahren I (Archaeen): Den jeweiligen Enzymklassen der extrahierten enzymatischen Funktionen der BRENDA werden Gensequenzen zugeordnet und gegen die UniProt-TrEMBL-Datenbank geblastet. Dies entspricht einer basalen Form der Enzymannotation. Der Ergebnismenge (E-Value geringer als 10⁻³, Identitätswert größer als 45%) an Genloci verschiedener Organismen werden die entsprechenden Domänen Archaea und Bacteria zugeordnet, sodass folgende Gruppierungen entstehen:

- (Ia) Archaeale und bakterielle Sequenzen mit hoher Homologie untereinander und zur enzymatischen Funktion der Suchsequenzen.
- (Ib) Archaeale und bakterielle Sequenzen mit hoher Homologie untereinander, jedoch nicht zur enzymatischen Funktion der Suchsequenzen.
- (IIa) Archaeale Sequenzen mit geringer Homologie zu den Sequenzen der Bakterien, jedoch einer Ähnlichkeit zur enzymatischen Funktion der Suchsequenzen im Rahmen der Schwellenwerte.
- (IIb) Archaeale Sequenzen mit geringer Homologie zu den Sequenzen der Bakterien und zur enzymatischen Funktion der Suchsequenzen außerhalb des Rahmens der Schwellenwerte.
- (IIIa) Bakterielle Sequenzen mit geringer Homologie zu den Sequenzen der Archaeen, jedoch einer Ähnlichkeit zur enzymatischen Funktion der Suchsequenzen im Rahmen der Schwellenwerte.
- (IIIb) Bakterielle Sequenzen mit geringer Homologie zu den Sequenzen der Archaeen und zur enzymatischen Funktion der Suchsequenzen außerhalb des Rahmens der Schwellenwerte.
- (IV) Sequenzen anderer Domänen als Archaea und Bacteria mit hoher Homologie untereinander und zur enzymatischen Funktion der Suchsequenzen.

Die Gruppen (Ib), (IIb), (IIIb) und (IV) werden in allen anschließenden Analysen zur jeweiligen Enzymklasse ausgeschlossen. Die Auswertung basiert auf (Ia), (IIa) und (IIIa). Die Quantität dieser einzelnen Gruppen schwankt je nach Ergebnis, bzw. Ursprung des Enzyms. Dementsprechend kann beispielsweise ein **archaeentypisches Enzym** eine große Anzahl zur Funktion homologer Sequenzen (IIa) aufweisen, während ebenso wenige bis gar keine Überschneidungen mit bakteriellen Sequenzen (Ia) vorliegen, wie bakterielle Gene in entsprechender Distanz, jedoch im Rahmen der Schwellenwerte, (IIIa) gefunden werden. Ist andererseits in diesem Beispiel die Ergebnisgruppe (Ia) relativ klein und werden demgegenüber Domänen-anteilig gleich viele bakterielle Gene verschiedener Organismen wie bei den Archaeen annotiert, ist eine mehrfach evolvierte domänentypische enzymatische Funktion zu vermuten. (Kapitel 3.4)

Die Zuordnung der recherchierten Metadaten (Kapitel 2.4) zu den Archaeen der Ergebnismenge (Ia) und (IIa), ermöglicht die Verknüpfung entsprechender Zusammenhänge mit der enzymatischen Funktion. (Kapitel 3.4)

Das Vorgehen nach Verfahren II

In einem zweiten Verfahren dienen die EnzymeDetector Ergebnisse zu allen analysierten Rhodobacteraceen und Archaeen als Grundlage für die Auswertung mit einem erstellten Python-Skript (Anhang 5) (Abbildung 2).

Die Rohannotationen der Enzyme zu einzelnen Genen aus verschiedenen Datenbanken bedarf einer Optimierung im Hinblick darauf, welche bekannte Funktion jene ist, die am wahrscheinlichsten erscheint. Dazu bietet sich eine Auswertung der EnzymeDetector Relevanzwerte für jeden einzelnen Genlocus nach folgenden Kriterien an [119]:

Die Kriterien einer optimalen Genlocusannotation:

- Annotationen mit einem Relevanzwert kleiner als ein Schwellenwert werden ignoriert.
- Nicht sequenzbasierte Annotationen werden nur über dem Schwellenwert akzeptiert.
- Annotationen mit einem Relevanzwert größer als ein Schwellenwert werden ohne weitere Bedingungen in die Ergebnismenge aufgenommen.
- Ist für einen Genlocus nur eine Annotation einer Enzymfunktion innerhalb des Schwellenwertbereichs bekannt, wird diese akzeptiert.
- Werden unterschiedliche Annotationen für einen Genlocus innerhalb des Schwellenwertbereichs angegeben:
 - Wird nur eine Annotation zugelassen, sollte sie einen Relevanzwert doppelt so hoch wie alle anderen aufweisen.
 - Wird nur eine Annotation zugelassen, sollte ihr BLAST-Ergebnis eine um 1/3 bessere Identität als alle anderen aufweisen.
 - Werden mehrere Annotationen zugelassen, die einen über die Hälfte besseren Relevanzwert als die restlichen aufweisen und deren BLAST-Ergebnis auf bekannte multifunktionale Enzyme in SwissProt verweist. Es werden nur Sets von Enzymfunktionen zugelassen, deren Positionen auf der Sequenz sich um maximal 8 Aminosäurepositionen überlappen (multifunktionale Enzyme).
 - Werden alle Annotationen im Schwellenwertbereich zugelassen, sollte keine Entscheidung nach obigen Kriterien möglich sein.

Als Schwellenwert für die obere Grenze wird ein Relevanzwert von 36 angesetzt. Dies akzeptiert Annotationen, die in manuell geprüften Datenbanken vorkommen und denen ein Genlocus zugewiesen ist, in allen Fällen. Für die untere Grenze wird der Schwellenwert angenommen, für den maximal 30% des kompletten Genoms eine enzymatische Funktion nachgewiesen bekommen [23]. Der minimale untere Schwellenwert ist jedoch mit einer Relevanz von 4 festgelegt. Dies ist dadurch bedingt, dass Fehlinterpretationen falscher Annotationen durch vereinzelte Datenbanken oder nur einzelner Verfahren ausgeschlossen werden sollen. Für die Auswertung werden immer Chromosomen- und Plasmidannotation als eine Einheit für die jeweiligen Organismen betrachtet. Diese so gefilterten und optimierten EnzymeDetector Ergebnisse resultieren in einer Matrix, die beschreibt, welche Enzyme in welchem Organismus vorkommen oder nicht. Des Weiteren kann mit der Information, welche Enzyme wiederum für welche Pathways essentiell notwendig sind (Kapitel 2.4 und Kapitel 3.2), diese Matrix auch auf Stoffwechselwegebene erstellt werden. Bildet man nun die

gesammelten Metadaten auf die zugehörigen Organismen und ihre zugeordneten Enzyme und Stoffwechselwege ab, lassen sich gehäufte Verbindungen erkennen und zur Interpretation herausgreifen (Kapitel 3.5 und Kapitel 3.6). Diese Abbildung der Daten auf die Metadatenkategorien erfolgt in Excel, während bedingte Farbformatierungen per Farbskala die Anteile der annotierten Organismen wiedergeben. Dadurch können auffällige Häufigkeiten einzelner Enzyme oder Stoffwechselwege einer mikrobiellen Eigenschaft erkannt werden (Anhang 1 und 2).

Weiterhin werden Enzyme selektiert, die in der Projektgruppe nicht annotiert sind und somit gehäuft zu „Lücken“ in den betrachteten Stoffwechselwegen führen. Sind alle benötigten restlichen Enzyme eines Pathways allerdings von der untersuchten Organismengruppe zu großen Anteilen annotiert und ist das betreffende Enzym obligatorisch im Reaktionsverlauf einzustufen, führt dies immer wieder zum Ausschluss aus den Ergebnissen der Matrix. Kann der Stoffwechselweg allerdings als essentiell für die phylogenetische Klade oder zumindest als nicht abwegig erkannt werden, können damit eventuell Rückschlüsse auf unbekannte und gruppeneigene Enzymfunktionen gemacht werden (Kapitel 3.7). Die Extraktion der angesprochenen Enzyme, die mangels obligatorischer Annotation Stoffwechselwege ausschließen, erfolgt mit dem erstellten Python-Skript (Anhang 5) zur Auswertung der Enzymzuweisungen und der Zuordnung zu Pathways.

Zur weiteren Analyse können aus den Enzymmatrizen in den Organismengruppen von Interesse häufig annotierte Enzyme, für die allerdings kein Pathway bekannt ist, Hinweise auf neue Stoffwechselwegansätze innerhalb der phylogenetischen Gruppe geben (Kapitel 3.8).

Die Analyse der zusammengetragenen Lücken und Enzyme ohne Zuordnung eines Stoffwechselweges wird mit Excel (Anhang 1 und 2) und den *Pathway*-Informationen der MetaCyc-Datenbank auf dem Stand 12/2015 bewerkstelligt, sollten sie jeweils zu größeren Anteilen in den untersuchten Organismengruppen auftreten.

Wie eingangs erwähnt, dient der Quervergleich von Organismenstämmen oder gar ganzen phylogenetischen Kladen dazu, Lücken der Annotationen ohne direkten Nachweis zu füllen. Diesem Ansatz geht die grundsätzliche Annahme voraus, dass der „Kern der Enzyme“, dies entspricht der Menge an Enzymen, die in allen eingeschlossenen Organismen vorkommt, entsprechend groß ist. Hierzu wird in der Arbeit der Begriff des *Core*-„Enzynoms“ in Anlehnung an die Genetik geprägt. [120] Diese Enzyme können jedoch in jedem Fall bei der Charakterisierung einer Klade eine Hilfestellung geben. Hingegen kann man die Menge aller verschiedenen codierten Enzyme einer phylogenetischen Gruppe als Maß für die Vielseitigkeit heranziehen. [120] Auch hierfür soll der Begriff des *Pan*-„Enzynoms“ abgeleitet aus der Genetik stehen (Kapitel 3.9). Die Auswertung der per Python-Skript aufgestellten Analysen (Anhang 5) der Enzymannotationen erfolgt in einer Excel-Datentabelle (Anhang 1 und 2), die das *Core*- und *Pan*-„Enzynom“ zusammenstellt.

Zum Vergleich dieser Ergebnisse mit den anerkannten phylogenetischen Verwandtschaftsverhältnissen auf Ebene der 16sRNA werden Binärvektoren zu allen Organismen bezogen auf alle in den untersuchten Rhodobacteraceen, bzw. Archaeen zugewiesenen Enzyme und Stoffwechselwege erzeugt (vorhanden 1 | nicht vorhanden 0). Die Distanzmatrix wird durch einfaches Zählen der

Unterschiede der jeweiligen Vergleichsvektoren erzeugt (Hamming-Distanz [102]). Als Außengruppe für das Rooting des resultierenden phylogenetischen Baumes wird ein idealer „Outgroup-Vektor“ erstellt, der alle betrachteten Enzyme bedient (Kapitel 2.3, Kapitel 3.10 und Kapitel 3.11).

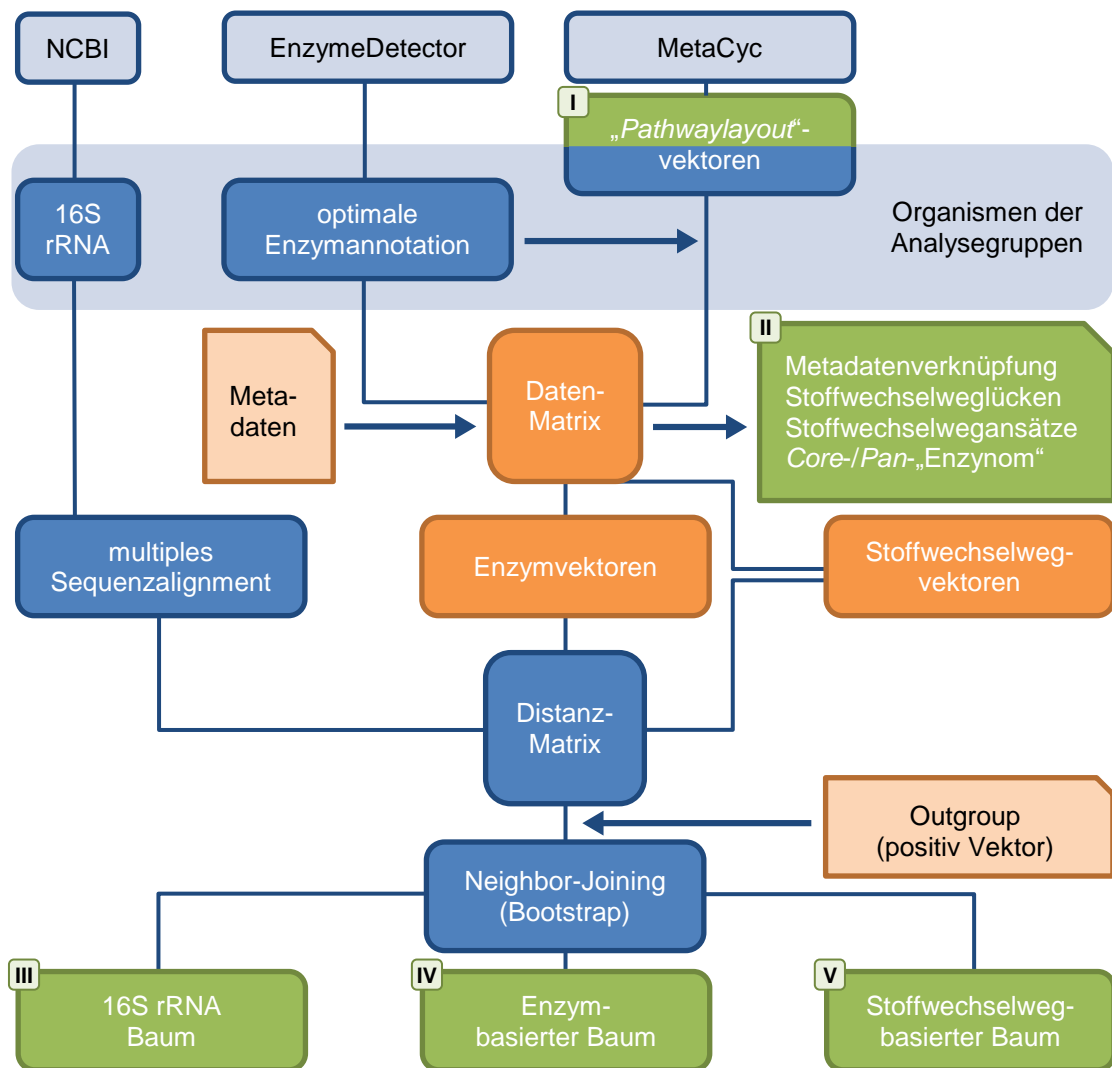


Abbildung 2: Schema Verfahren II (Analysegruppen): Im Zuge des zweiten Verfahrens werden, der jeweiligen Analysegruppe entsprechend, die 16SrRNA Sequenzen aus dem NCBI und die EnzymeDetector-Daten der Organismen extrahiert. Essentiell für die spätere Zuordnung von Stoffwechselwegen zu einem Enzympool eines Organismus ist die Erstellung von „Pathwaylayout“-Vektoren, um obligatorische enzymatische Schlüsselfunktionen vorauszusetzen und alternative Teilwege zu erkennen. (I), Kapitel 3.2) Diese Zuordnungen auf Grundlage der optimalen Enzymannotationen (siehe: Die Kriterien einer optimalen Genlocusannotation, Kapitel 2.9) aller Gene resultieren in einer Datenmatrix, die sowohl den Enzympool, als auch verfügbare Stoffwechselwege der einzelnen Organismen beschreibt. Diese wird mit den recherchierten Metadaten verknüpft (II) und zur Interpretation hinsichtlich auffälliger Zusammenhänge analysiert. (Kapitel 3.5, Kapitel 3.6) Weiterhin erfolgen Rückschlüsse auf eventuelle gruppeneigene enzymatische Funktionen durch wiederholt auftretende Annotationslücken in Stoffwechselwegen. (Kapitel 3.7) Ebenso werden häufig annotierte Enzymklassen ohne einen zugeordneten Pathway bezüglich eines funktionellen metabolischen Zusammenhangs diskutiert (Kapitel 3.8) und die Gesamtheit des Enzympools der Analysegruppe analysiert. (Core-/Pan-„Enzymom“, Kapitel 3.9)

Zum Vergleich mit den anerkannten phylogenetischen Verwandtschaftsverhältnissen der Analysegruppen auf Basis der 16S rRNA, werden diese durch die Berechnung eines multiplen Sequenzalignments in eine Distanzmatrix überführt. Diese dient als Grundlage zur Berechnung des phylogenetischen Baums (III) mit Hilfe des Neighbor-Joining Algorithmus (100 Bootstrap-Replikate). (Kapitel 2.3) Als Außengruppe werden hier bekannte Organismen mit entsprechender Distanz zur Analysegruppe für die Durchführung eines „Midpoint-Rootings“ gewählt. Nach dem gleichen Verfahren werden phylogenetische Bäume auf Grundlage der annotierten Enzyme (IV) und der zugeordneten Stoffwechselwege (V) berechnet und im Hinblick auf Differenzen analysiert. (Kapitel 3.10, Kapitel 3.11) Dazu wird die Datenmatrix aus (II) in binäre Vektoren mit Bezug auf alle erkannten Enzymklassen und Stoffwechselwege der Analysegruppe überführt und die Distanzmatrix der Organismen bestimmt (Hamming-Distanz). Diese Distanz wird ausgehend von einem manuell erstellten Positivvektor berechnet, der alle Kriterien (Enzyme, bzw. Stoffwechselwege) zugesprochen bekommt und infolgedessen als Outgroup dient. (Kapitel 2.3)

3 Ergebnisse und Diskussion

3.1 Die Verbesserungen und Erweiterungen des EnzymeDetector

Der EnzymeDetector stellt durch die Aggregation verschiedener Annotationsquellen und eigener Vorhersagealgorithmen (Kapitel 1.6) eine Möglichkeit dar, die Problematik der bioinformatischen Enzymvorhersage (Kapitel 1.3) einzuschränken und die wahrscheinlichste enzymatische Funktion eines Gens zu bestimmen. Im folgenden Kapitel werden die einzelnen Optimierungsschritte des Programms beschrieben, die die ursprüngliche Version des EnzymeDetector grundlegend verbessert und erweitert haben.

Algorithmische und logische Anpassungen

Durch den modularen Aufbau des EnzymeDetector, der Vorberechnung aller Zwischendatenbanken der Annotationsquellen und durch weitere Korrekturen und Anpassungen des Quellcodes ist es gelungen, die benötigte Bearbeitungszeit eines Updatezyklus auf etwa 6-7 Wochen zu reduzieren. Dazu zählt auch die sinnvolle SQL-Indexierung der Datenbanktabellen, wodurch eine weitere Reduzierung der pro Organismus benötigten Zeit für das Zusammentragen aller Annotationsquellen und -verfahren erreicht wird. Vergleichswerte eines vollständigen Updatezyklus der ursprünglichen Version des EnzymeDetector liegen nicht vor, da auf Grund der enormen Rechenzeit nur einzelne Organismengruppen bearbeitet wurden.

Dabei werden pro Update alle vollständig geschlossenen Genomassemblies des NCBI ebenso automatisiert eingebunden, wie alle benötigten Flatfiles der Quellressourcen. Die oben angegebene Rechenzeit bezieht sich auf einen Updatezyklus inklusive einer kompletten BLAST-, Pfam- und BrEPS-Analyse der zurzeit etwa 3000 Genome im finalen Status (Chromosomen und eventuelle Plasmide eines Organismus). In zukünftigen Updates kann gewählt werden, ob bereits vorhandene Gensequenzen nur erneut mit den Quellen auf aktuellem Stand abgeglichen werden sollen, oder ob auch der eigens durchgeführte BLAST, sowie die BrEPS- und Pfam-Analyse auf einen erneuerten Stand gebracht werden sollen. Dies setzt ein nennenswertes Update der entsprechenden Datenbank voraus. Sind folglich zum Beispiel nur sehr wenige neue Enzymannotationen in SwissProt zwischen zwei Updates hinzugekommen und ist der Organismus dem EnzymeDetector bereits bekannt, kann wahlweise der zeitkritische Schritt des BLASTs übersprungen und die bereits vorhandenen Daten genutzt werden.

Optimierungen der Bearbeitungszeit eines Updatezyklus

Vergleicht man die übernommene und die hier erarbeitete Version des EnzymeDetector am Beispiel eines Durchschnittsorganismus mit 3000 Genen, wie in Abbildung 3 zu sehen, so lässt sich eine Reduzierung der benötigten Zeit

für das Zusammentragen der Annotationsquellen auf etwa 1/8 der ursprünglich erforderlichen Zeit feststellen. Die übernommene Version des EnzymeDetector benötigt etwa 74 Stunden Gesamtberechnungszeit pro angesprochenem Organismus. Demzufolge würden für einen vollständigen Updatezyklus, der alle heute verfügbaren geschlossenen Genome einschließt, etwas mehr als 25 Jahre vergehen. Der zeitkritische Schritt des BLASTs kann vor allem durch die Cluster-Implementierung von im Schnitt benötigten 72 Stunden auf 15 Minuten verringert werden. Ebenso ist die Verminderung der beanspruchten Zeit für die BrEPS-Analyse auf 1/15 zum Großteil durch die Anpassung von MySQL-Statements, SQL-Indizierungen und Tabellenstrukturen zu vermerken. Alles in Allem kann die Bearbeitungszeit von zunächst circa 3 Tagen für einen 3000 Gene umfassenden Organismus, trotz zusätzlicher Annotationsquellen und durchgeführter Pfam-HMM-Suche, auf etwas mehr als 20 Minuten reduziert werden. Somit ist ein halbjährlicher Updatezyklus planbar und praktikabel.

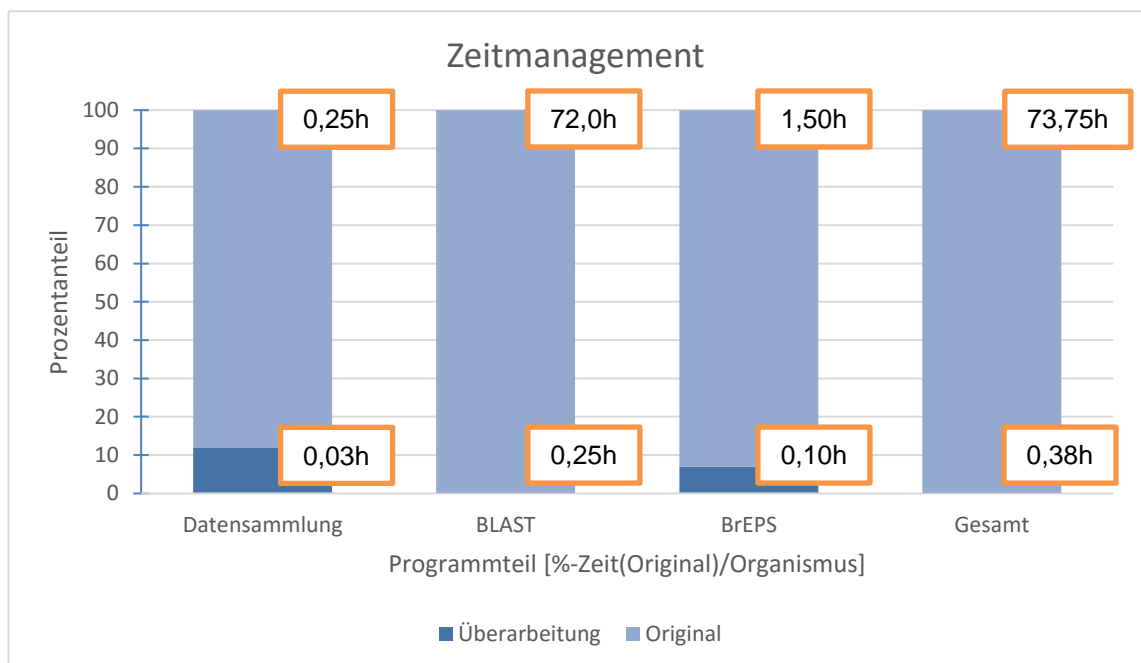


Abbildung 3: Vergleich des Zeitmanagements der ursprünglichen und der überarbeiteten Version des EnzymeDetector am Beispiel eines Durchschnittsorganismus mit 3000 Genen.

Werden in der ursprünglichen Version noch die sogenannten vom NCBI eingeführten „Gene-Identifizier“ genutzt, um ein Mapping der Annotationen auf einzelne Gene über alle Datenbanken hinweg zu ermöglichen, wird dies nun über den Genlocus-Tag gewährleistet. Dies hat den Vorteil, auch vom NCBI völlig unabhängige Daten einbinden zu können, ohne zum NCBI mappen zu müssen. Ist aktuell beim *National Center for Biotechnology Information* die Rede von einem Gene-Identifizier, so ist dieser nur auf die Sequenz bezogen eindeutig. Eine auf Gene bezogene Eindeutigkeit kann nur über den entsprechenden LocusTag erlangt werden. [98]

Einbindung zusätzlicher Annotationsquellen und Erweiterungen

Die Einbindung der zusätzlichen Annotationsquellen PATRIC, den KEGG Orthologie-Daten und das Zuordnen von Pfam-Proteinfamilien (enzymbasiert) (Kapitel 1.9, Kapitel 2.4) geben weitere Hinweise auf entsprechende oder andere Enzymzuweisungen zu den Genloci verschiedener Organismen.

Der eigene Annotationsalgorithmus mit Hilfe des BLASTs gegen die SwissProt-Datenbank (Kapitel 1.6) wurde in zweierlei Hinsicht verbessert. Einerseits kann durch das Speichern der genauen Positionsangaben der Enzymsequenztreffer auf der Suchsequenz nun abgeschätzt werden, ob eventuell mehrere Enzymfunktionen vorliegen können, ohne den gleichen Genabschnitt zu translatieren. Ist hingegen die Treffersequenz in SwissProt schon bifunktional annotiert, kann mittels der Sequenzlänge geprüft werden, ob sich beide Enzymfunktionen auch im zu annotierenden Gen unterbringen lassen. Weist die Suchsequenz nicht einmal halb so viele Aminosäuren wie ein bifunktional annotiertes Gen auf, so wird höchstwahrscheinlich auch nur eine der beiden Funktionen auf diesem Gen codiert sein. Diese Information hingegen geht beim ursprünglichen BLAST-Algorithmus verloren. Auf Grund des möglicherweise guten *E-Values* (Kapitel 2.2) würden beide Enzymfunktionen zugewiesen.

Auf der anderen Seite steht die Problematik, dass die Grundanlage des Algorithmus versucht, in jedes Gen eine enzymatische Funktion zu interpretieren. Da Untersuchungen an Referenzorganismen allerdings zeigen, dass schätzungsweise nur etwa maximal 30% des Genoms enzymcodierend sind, ist hiervon keinesfalls auszugehen. [23] Der BLAST bietet für diese Entscheidung nun eine Hilfestellung an. Zählen nicht enzymatische Gene in der SwissProt-Datenbank ebenfalls zu der Ergebnismenge des BLASTs, wird die beste (kleinster *E-Value* und größte Identität) und längste (längster zusammenhängender Trefferabschnitt von Start bis Ende) Übereinstimmung vermerkt, insofern sie damit als wahrscheinlicher als alle Enzymannotationen eingestuft werden kann. Auf diese Weise können zwangsinterpretierte Annotationen verworfen werden.

Die Beschränkung der BLAST-Annotationen auf jene, die durch mindestens 5 verschiedene Treffersequenzen die gleiche Enzymnummer zugewiesen bekommen, wird übernommen (1. Qualitätsstufe) (Kapitel 1.6). Da das Überschreiten dieses Schwellenwertes allerdings direkt davon abhängig ist, wie oft ein Enzym schon annotiert ist, bzw. ob es überhaupt schon mehr als 5 Zuordnungen gibt, könnte ein Verwerfen dieser Annotationen zum Informationsverlust führen. Infolgedessen wird die 2. Qualitätsstufe in der Ergebnismenge nun neben der strikt gefilterten 1. Qualitätsstufe mit einer geringeren Bewertung gesondert mit aufgenommen.

Anpassungen des Bewertungsschemas des Konfidenzwerts

Das Scoring-Schema des EnzymeDetector soll durch die Qualitätseinstufung der Annotationsquellen dabei helfen, die wahrscheinlichste Enzymfunktion zu deklarieren.

Vorliegende statistische Untersuchungen zeigen, dass eine Aggregation durch einfaches „Zusammenzählen“ kongruenter Annotationen ohne Gewichtung, jedoch mit der Unterscheidung zwischen automatisiert generierten und experimentell, bzw. in der Literatur nachgewiesenen Annotationen, ausreichen könnte. [121] Dahingehend wurde das Schema für diese Arbeit angepasst, bzw. erweitert. Ein simples Aufsummieren der entsprechenden einzelnen Annotationsquellen entspricht einer Gleichgewichtung von NCBI, KEGG, KEGG Orthologie, PATRIC, BLAST, BREPS und Pfam. Allerdings muss dieser Umstellung eine statistische Analyse über einen großen Referenzbereich vorhergehen. Dies ist nicht im Rahmen dieser Arbeit erfolgt. Demgemäß findet das Bewertungsschema aus Tabelle 4 hier Anwendung, infolgedessen Konfidenzwerte zwischen 1 und 163 bei maximaler Kongruenz aller einfließenden Annotationsquellen vergeben werden.

Eine Abstufung der Zuverlässigkeit der durch das Textmining der AMENDA-Datenbank (Kapitel 2.4) automatisiert festgestellten Verknüpfungen von Enzymen zu Organismen ermöglicht die Einordnung in verschiedene Qualitätskategorien. Dabei wird eine Bewertung vorgenommen, inwieweit die Bezeichnungen von Enzymen direkt in Konkurrenz mit den Organismennamen auftreten. Zum Beispiel erhält eine enzymatische Funktion unmittelbar gefolgt von einer Spezies im Titel der Publikation die Höchstbewertung. [92] Diese Information wird nun durch die daraus resultierende Differenzierung des vergebenen Konfidenzwertes wiedergegeben. Weiterhin wird unterschieden, welche Annotationen sich direkt auf den besagten Stamm der untersuchten Spezies beziehen, oder ob nur eine Aussage auf Ebene des Genus gemacht werden kann. Der eingangs erläuterte Mehrwert manuell geprüfter und experimentell nachgewiesener Annotationsdatenbanken (BRENDA, SwissProt) schlägt sich, wie auch in der übernommen EnzymeDetector Version, in einem vielfach höheren Score von 50 im Vergleich zu anderen Quellen nieder. Zur erweiterten Einordnung der BRENDA-Annotationen im Hinblick auf einen Organismenstamm von Interesse wird nun nicht mehr zwangsweise wie im originalen Algorithmus des EnzymeDetector ohne Berücksichtigung von Annotationsdetails, wie Organismusstamm, Genlocus und dem Basieren auf Sequenzzuordnungen, der maximal erreichbare Score von 50 vergeben. Die Bewertungsbereiche der E-Value-Werte der neuen zweiten Qualitätsstufe des BLASTs und der Pfam-Annotation werden ebenso aus der übernommenen Version des EnzymeDetector übertragen. Dem nicht Überschreiten der mindestens geforderten fünf verschiedenen Treffersequenzen für eine Annotation im Fall der zweiten BLAST Qualitätsstufe, wird mit einer Halbierung des Scores Rechnung getragen. Da Proteinfamilien oft für mehrere Enzyme mit ähnlicher Funktion stehen und diese Annotationen dementsprechend eher als Hinweis verstanden und nicht zur expliziten Unterscheidung eng verwandter Enzyme genutzt werden können, wird der Pfam-Bewertung demgemäß eine

3 Ergebnisse und Diskussion

Abstufung von 1-4 zugewiesen. Den zusätzlich genutzten Datenbanken der KEGG Orthologie und PATRIC wird ein Score von 5 zugewiesen. Dies ermöglicht die Abgrenzung zu den bereits in der übernommenen Version des EnzymeDetector enthaltenen automatisiert generierten Annotationsdatenbanken mit einer Bewertung von 3 (NCBI, KEGG) und drückt zugleich die zu erwartende höhere Genauigkeit und Sensitivität der angewendeten Verfahren zur Annotation [85,90] aus.

Tabelle 4: Scoring-Schema des EnzymeDetector. Die Tabelle listet die einzelnen Bewertungen des überarbeiteten EnzymeDetector auf. Der maximal erreichbare Score einer Quelle entspricht dem Konfidenzwert. Dabei werden die einzelnen Abstufungen der Pfam, BLAST, BrEPS, AMENDA und BRENDA Konfidenzwerte gesondert aufgeführt.

Quelle	Konfidenzwert
NCBI, KEGG	3
KEGG-Orthologie, PATRIC	5
Pfam	E-Value < 10 ⁻¹²⁰ : 4 E-Value < 10 ⁻⁸⁰ : 3 E-Value < 10 ⁻⁴⁰ : 2 E-Value > 10 ⁻⁴⁰ : 1
BLAST [1.Qualitätsstufe]	E-Value < 10 ⁻¹²⁰ : 8 E-Value < 10 ⁻⁸⁰ : 6 E-Value < 10 ⁻⁴⁰ : 4 E-Value > 10 ⁻⁴⁰ : 2
BLAST [2.Qualitätsstufe]	E-Value < 10 ⁻¹²⁰ : 4 E-Value < 10 ⁻⁸⁰ : 3 E-Value < 10 ⁻⁴⁰ : 2 E-Value > 10 ⁻⁴⁰ : 1
BrEPS	1-10, nach <i>Correctness</i> des Sequenzmusters
SwissProt	50
AMENDA	<u>Organismus Stamm:</u> Rating „++++“: 25 Rating „+++“: 20 Rating „++“: 10 Rating „+“: 5 <u>Organismus:</u> Rating „++++“: 15 Rating „+++“: 10 Rating „++“: 5 Rating „+“: 2
BRENDA	<u>Organismus Stamm:</u> 50,45,40 <u>Organismus:</u> 50,10,5 [Abstufungen siehe Tabelle 4]

Differenzierung der Enzymannotationen der BRENDA

Die Tabelle 5 verdeutlicht die angewendete Differenzierung der BRENDA-Annotationen. Die Bewertung wird in die Anteile 10% für die Annotation an sich, 60% für Annotationen, die direkt dem Stamm zugeordnet werden können, bzw. 20% für Annotationen, die direkt einem Gen des Stammes zuzuordnen sind, aufgegliedert. Die übrigen 5% oder 5 Bewertungspunkte differenzieren darin, ob es sich um sequenzbasierte Annotationen handelt oder nicht. An dieser Stelle ist

festzuhalten, dass nur ein Gesamtscore von 50 für Annotationen zu einem Gen des gesuchten Stammes mit der Verknüpfung zu einer Sequenz vergeben wird. Dies ist auch möglich, sollte bei der Genus spezifischen Suche ein hinterlegtes Gen explizit dem Organismus zugeordnet werden können (Tabelle 5, Genussuche, Score 50, rot markiert). Liegt keine Sequenz vor, werden die entsprechenden 10% abgezogen und in der Ergebnistabelle mit „not sequenced based“ vermerkt.

Nicht sequenzbasierte, jedoch stammbezogene Suchergebnisse werden auch ohne vorliegendem Genbezug dem Organismus-Genom zugeordnet (Tabelle 5, Stammsuche, Score 45, rot markiert). Liegt hingegen eine Sequenz vor und kann diese aber nicht über den Locus dem betrachteten Organismus zugeordnet werden, wird eine Gesamtbewertung von 40 wiedergegeben. Der Abzug spiegelt die Vermutung wieder, dass derartige Annotationen aus der gleichen Spezies stammen, jedoch auf anderen Organismusstämmen beruhen. Ergebnisse, die nur durch die Suche des Genus hervorgebracht werden, können, insofern sie ein Gen des Stammes verlinken, ebenfalls die Maximalpunktzahl erreichen. Die Abstufungen von 10, bzw. 5 als Gesamtbewertung resultieren folglich noch aus sequenzbasierten Annotationen anderer Stämme der Spezies und Zuweisungen zum Genus ohne genetischem Nachweis.

Tabelle 5: Das BRENDA-Scoring-Schema. Der vom EnzymeDetector maximal zu vergebende Konfidenzwert für die BRENDA-Annotationen wird in einzelne Prozentklassen unterteilt. Im einfachsten Fall erhält eine Annotation folglich ohne Verweise auf den gesuchten Organismusstamm, einen Genlocus und eine verknüpfte Sequenz, 10% des maximalen Konfidenzwertes von 50. Weitere Informationen summieren die Bewertung einer Annotation weiter auf. Dementsprechend werden 60% des Maximalwertes addiert, sollten die Annotationen mit der gesuchten Stammbezeichnung hinterlegt sein. Ist das hinterlegte Gen mit der Genlocus-Bezeichnung dem Genom des entsprechenden Organismus zuzuordnen, werden weitere 20% von 50 aufgeschlagen. 10%, bzw. fünf Konfidenzwertpunkte kommen hinzu, insofern Gensequenzdaten vorliegen. Annotationen ohne Sequenzinformationen, jedoch mit Bezug zum Organismusstamm, besitzen in logischer Konsequenz auch keinen hinterlegten Genlocus. Um dennoch den Bezug auf einen expliziten Stamm aufzuwerten, wird derartigen Zuordnungen der Genombezug automatisch zugesprochen. (rot, Score 45) In der BRENDA existieren Annotationen aus publizierten Quellen ohne Stammesbezug, allerdings mit der Angabe eines Genlocus. Ist wiederum dieses Gen dem Organismus von Interesse zuzuordnen und eine Sequenz ermittelbar, werden ebenfalls die 60% für die Stamminformation aufgeschlagen, sodass der maximale Konfidenzwert von 50 vergeben wird. (rot, Score 50)

	Score max. 50	Annotation 10%: 5	Strain-Annotation 60%: 30	Organismus-Genom 20%: 10	sequenzbasiert 10%: 5
Stamm- Suche	50	X	X	X	X
	45	X	X	(„-“) → X	-
	40	X	X	-	X
Genus- Suche	50	X	(„-“) → X	X	X
	10	X	-	-	X
	5	X	-	-	-

Das neue Webinterface des EnzymeDetector

Zur Weitergabe und zum Überblick der EnzymeDetector-Daten bietet das Webinterface nun mit einem gruppierten und sortierten Glossar der Organismen (Abbildung 4) einen einfachen Einstieg.

The screenshot displays the EnzymeDetector web interface. At the top, there are logos for Technische Universität Braunschweig, a diamond-shaped logo, and the BRENDA logo (The Comprehensive Enzyme Information System). Below these is the text 'EnzymeDetector' and 'An integrated enzyme function prediction tool and database.' with links for 'Introduction', 'Select:', and 'Documentation'. The main section is titled 'Organism' and features a navigation bar with buttons for 'ALL', '0-9', and letters 'A' through 'Z'. The letter 'D' is highlighted. Below the navigation bar, a list of organisms is displayed, including Dactylococcopsis salina PCC 8305, Dechloromonas aromatica RCB, Dechlorosoma suillum PS, Deferribacter desulfuricans SSM1, Dehalobacter sp CF, Dehalobacter sp DCA, Dehalococcoides ethenogenes 195, Dehalococcoides mccartyi BTF08, Dehalococcoides mccartyi DCMB5, Dehalococcoides sp BAV1, Dehalococcoides sp CBDB1, Dehalococcoides sp GT, Dehalococcoides sp VS, Dehalogenimonas lykanthroporepellens BLDC9, Deinococcus deserti VCD115, Deinococcus geothermalis DSM 11300, Deinococcus gobiensis J0, Deinococcus maricopenensis DSM 21211, Deinococcus perardilitoris DSM 19664, Deinococcus proteolyticus MRP, and Deinococcus radiodurans R1.

Abbildung 4: EnzymeDetector Organismenglossar. Alle Organismen mit Annotationsinformationen des EnzymeDetector werden alphabetisch sortiert in einem Glossar zu Auswahl angeboten. Eine Schnellauswahl der gruppierten Anfangsbuchstaben ist im oberen Bereich möglich.

Die Entscheidung über einzustellende Schwellenwerte wird dem Nutzer durch eine Voreinstellung abgenommen. Dabei wird wieder berücksichtigt, dass zumeist etwa 30% eines Genoms maximal für enzymatische Funktionen codieren. Allerdings kann hier ebenso eine freie Wahl getroffen werden wie bei der Einstellung der Relevanzwerte der einzelnen Annotationsquellen. Nach Genloci, „Gene-Identifizier“, UniProtID, EC-Nummer, Enzymname, Translationsrichtung und Pfam-Accession können in einem freien Eingabefeld entsprechende Einträge sofort gesucht werden. Ebenso können alle genannten Spalten mit einem Smartfilter unter der Tabelle eingeschränkt werden. Dieser Smartfilter übernimmt direkt ohne erneute Ladezeiten das Filterergebnis in die schon angezeigte Tabelle. Der komplette Datensatz oder nur markierte Einträge stehen dem Nutzer zum Druck, für die Zwischenablage oder als Download im CSV-, bzw. PDF-Format zur Verfügung. (Abbildung 5) Eine Statistikübersicht bietet zu jedem Organismus die Möglichkeit, den Überblick über die vorliegenden Annotationen zu erhalten. So erfolgt je nach erreichtem Gesamt-Score die Einordnung in Relevanzgruppen. Weiterhin wird dargestellt, wie viele Enzyme im Genom einmalig codiert sind, oder ob eine hohe Replikation vorliegt. Ferner werden die Annotationsquellen nach der Anzahl ihrer Zuweisungen unter Berücksichtigung der gewählten Schwellenwerte aufgelistet. Eine Detailansicht

der dem Organismus verfügbaren Enzyme und wiederum der daraus resultierenden möglichen Pathways, gibt einen ersten Aufschluss darüber, welche Stoffwechselwege komplett oder nur zu bestimmten Anteilen abgedeckt, bzw. welche Enzyme nicht annotiert sind. Des Weiteren kann der Vergleich zu einem beliebig wählbaren anderen Organismus in der Datenbank des EnzymeDetector erstellt werden. So können charakteristische Enzyme von zwei Organismen genauso herausgestellt werden, wie die Schnittmenge beider.

Der Frage der Enzymverteilung in den Domänen der Bakterien und Archaeen kann durch die Auswahl des Enzymnamens, bzw. der EC Nummer auf der Seite des Enzymglossars nachgegangen werden. Die entsprechenden Annotationen im gewählten Relevanzbereich werden hier verknüpft mit den Genomen, taxonomischen IDs und den Genom-Accessions übersichtlich dargestellt.

Organism:

Dinoroseobacter shibae DFL 12
Chromosome: NC_009952
Plasmid(s): NC_009959: pDSH03, NC_009958: pDSH04, NC_009956: pDSH02, NC_009955: pDSH01, NC_009957: pDSH03

Total genes: 4192
Enzyme coding genes: 1334 (32% of genome)
Relevance-CutOff: 6
* small values increase loading time.

Relevances: * load/blank for default scores.
BRENDA: AMENDA: SwissProt:
NCBI: KEGG: KEGG(Orthology):
BREP5: PFAM: BLAST:
PATRIC:

Create Statistics
View Pathways
Compare Organism

Show entries

Chr/P	Locus	GI	UniProt ID	EC-Number	Relevance	recommended Name	Gene-Start	Gene-Stop	Gene-Direction	BLAST-eValue	BLAST-Identity
chr	Dsh_0268	159042824	AELM40	2.7.7.6	88	DNA-directed RNA polymerase	284787	288923	+	0.0	100.00
chr	Dsh_1593	159044142	AELKX1	2.3.1.46	88	homoserine O-succinyl transferase	1659426	1660343	+	4.66e-213	100.00
chr	Dsh_1642	159044191	AELK20	6.1.1.7	88	Alanine-DNA ligase	1701885	1704548	-	0.0	100.00
chr	Dsh_2742	159045282	AELN4	6.3.4.4	88	Adenylsuccinate synthase	2903816	2905111	-	6.08e-292	100.00
chr	Dsh_0031	159042587	AELJ74	2.7.1.130	87	tetracyl diaccharide 4-kinase	34226	35239	-	9.95e-228	100.00
chr	Dsh_0188	159042744	AELLC7	1.1.1.2	87	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	189644	190600	-	8.8e-210	100.00
chr	Dsh_0269	159042825	AELM41	2.7.7.6	87	DNA-directed RNA polymerase	289021	273268	+	0.0	100.00
chr	Dsh_0432	159042987	AELN36	3.5.3.6	87	arginase deiminase	423677	424906	+	8.15e-287	100.00
chr	Dsh_0977	159043526	AELS12	2.5.1.7	87	UDP-N-acetylglucosamine 1-carboxyvinyl transferase	998735	1000003	+	2.38e-273	100.00
chr	Dsh_1022	159043571	AELS80	2.5.1.19	87	3-phosphoshikimate 1-carboxyvinyl transferase	1057354	1058706	+	5.16e-302	100.00
chr	Dsh_1805	159044354	AELMK2	6.1.1.17	87	Glutamate-DNA ligase	1878120	1879535	+	0.0	100.00
chr	Dsh_2705	159045245	AELI7	4.1.1.37	87	Uroporphyrinogen decarboxylase	2865197	2866231	+	2.11e-229	100.00
chr	Dsh_3275	159043815	AELMT3	4.2.3.5	87	chorismate synthase	3439739	3440842	-	9.2e-250	100.00
chr	Dsh_0023	159042579	AELJ6	3.1.26.4	86	ribonuclease H	26998	27618	+	6.22e-133	100.00
chr	Dsh_0083	159042639	AELK29	4.2.1.33	86	3-isopropylmalate dehydratase	82492	83097	-	3.74e-133	100.00
chr	Dsh_0324	159042880	AELM96	2.7.2.8	86	acetylglutamate kinase	311973	312836	-	3.94e-183	100.00
chr	Dsh_0442	159042997	AELN46	3.6.3.14	86	H ⁺ -transporting two-sector ATPase	431227	432756	+	0.0	100.00
chr	Dsh_0794	159043346	AELQ26	4.2.1.19	86	imidazoleglycerol-phosphate dehydratase	801568	802155	+	1.51e-124	100.00
chr	Dsh_0922	159043474	AELRW0	2.7.4.9	86	dTMP kinase	946178	946801	-	8.02e-135	100.00
chr	Dsh_1394	159043943	AELJ74	2.3.2.6	86	leucyltransferase	1446363	1447034	-	9.15e-151	100.00

Showing 1 to 20 of 1,988 entries

Previous 2 3 4 5 ... 100 Next

Abbildung 5: Neues Webinterface des EnzymeDetector. (Ansicht der Annotationen zu *Dinoroseobacter shibae* DFL 12.) Informationen zum ausgewählten Genom (Genom-Accessions) werden im oberen Bereich links dargestellt. Rechts daneben können sowohl der Schwellenwert für den minimalen Konfidenzwert der anzuzeigenden Annotationen, als auch die einzelnen Scores der Quelldatenbanken eingesehen und angepasst werden. Die Schaltflächen für weitere Analysen mittels einer Statistikübersicht, der Zuordnung von Stoffwechselwegen und einem Genomvergleich befinden sich im Bereich rechts oben. Die Übersicht aller aufgeführten Enzymzuordnungen der einzelnen Genloci kann mittels einem Smartfilter über und unter der Tabelle mit den entsprechenden Suchbegriffen gefiltert werden.

Zusammenfassung

Die Erweiterungen und Anpassungen des EnzymeDetector ermöglichen eine Automatisierung der Erstellung einer möglichst optimalen und umfassenden Genannotation (Kapitel 1.3 und Kapitel 2.9) hoher Konfidenz von einer Vielzahl von Organismen. Dies findet in den folgenden Analysen der ausgewählten Organismengruppen Anwendung (Kapitel 3.5) und bewahrt vor Fehlinterpretationen nicht korrekter Genannotationen.

3.2 Das „Pathwaylayout“ mit MetaCyc

MetaCyc (Kapitel 1.9) beschreibt 2790 Stoffwechselwege von sehr unterschiedlicher Größe und Struktur. Die zugehörige Angabe des sogenannten „Pathwaylayouts“ in Form von Reaktionsketten der Substrate, Zwischen- und Endprodukte, ermöglicht eine automatisierte Unterscheidung beteiligter enzymatischer Schlüsselfunktionen, für die dieser Stoffwechselweg keine Alternativen vorsieht, von Enzymen, die über unterschiedliche Reaktionen dennoch zu konformen Produkten führen. Das hierfür erstellte Programm ist essentiell und von hoher Bedeutung für die Auswertung auf Ebene der Stoffwechselwege. Infolgedessen werden explizit nur Pathways den Organismen basierend auf ihrem Enzympool zugeordnet, die ihr Metabolismus funktionell bedienen kann.

Aus der MetaCyc-Analyse werden 84 Stoffwechselwege ohne Enzymbeteiligung und 1144 auf Grund von Reaktionsschritten mit unvollständig kategorisierten Enzymen ausgeschlossen. Ebenfalls können drei Stoffwechselwege vom erstellten Programm nicht aufgelöst werden. Dabei handelt es sich um die Folat Transformation I (MetaCyc: PWY-2201), dem „Superpathway“ des Alantoin Abbaus bei Pflanzen (MetaCyc: URDEGR-PWY) und die N¹⁰-Formyl-Tetrahydrofolat-Biosynthese (MetaCyc: 1CMET2-PWY). Die starke Vernetzung und das mehrfache Vorkommen ineinandergreifender Zyklen dieser sogenannten „Superpathways“ werden vom erstellten Programm nicht abgebildet, so dass keine Alternativvektoren erkannt werden.

Im Ergebnis stehen 1560 Pathways für die Auswertung der Organismengruppen zur Verfügung. Zu diesen ist jeweils ein vom Programm erzeugter Vektor bekannt, der obligatorische Enzymfunktionen ohne einen möglichen Ersatz beschreibt. Dazu werden Alternativvektoren, wenn möglich, gelistet. Diese bilden Seitenäste, bzw. zueinander alternative Reaktionsschritte ab, von denen nur jeweils ein Weg benötigt wird, um das Endprodukt des Stoffwechselweges zu bilden und somit den gesamten Pathway dem Organismus zusprechen zu können.

Bei der Auswertung der Rhodobacteraceen und der Domäne der Archaeen werden insgesamt 2376 verschiedene Enzyme zugeordnet. Daraus resultieren 655 verschiedene Pathways, die in der einen oder anderen Organismengruppe in entsprechenden Anteilen auf Grund der beteiligten Enzymfunktionen ablaufen können. 905 Stoffwechselwege werden in beiden Gruppen ausgeschlossen, da verbindlich benötigte Enzyme oder Alternativfunktionen in keinem analysierten Organismus vorhanden sind. Darunter befinden sich meist Stoffwechselwege, die der Domäne der Eukaryota zugesprochen werden oder spezieller bei Taxa wie den Metazoa oder den Viridiplantae zu erwarten sind [21].

Abbildung 6 zeigt die Bildung der obligatorischen und Alternativvektoren am Beispiel des TCA Zyklus. Gibt MetaCyc alternative Enzyme mit gleichen Reaktionspartnern und Cofaktoren wie im Falle der *citrate synthase* (EC 2.3.3.1 und EC 2.3.3.16) direkt an, gehen diese auch entsprechend so in den obligatorischen Vektor mit ein. Die Reaktion von (S)-Malat zu Oxalacetat wird zumeist nur bei Eukaryoten durch eine NAD-abhängige Dehydrogenase (EC

1.1.1.37) katalysiert. Prokaryoten hingegen weisen dafür oft eine Chinon-abhängige Alternative auf (EC 1.1.5.4) [21].

Die Alternativvektoren müssen sich nicht zwangsweise nur über einen Reaktionsschritt, wie in diesem Beispiel gezeigt, erstrecken. Die Wege können auch nach mehreren Zwischenschritten wieder zusammenführen und als Alternativen erkannt werden.

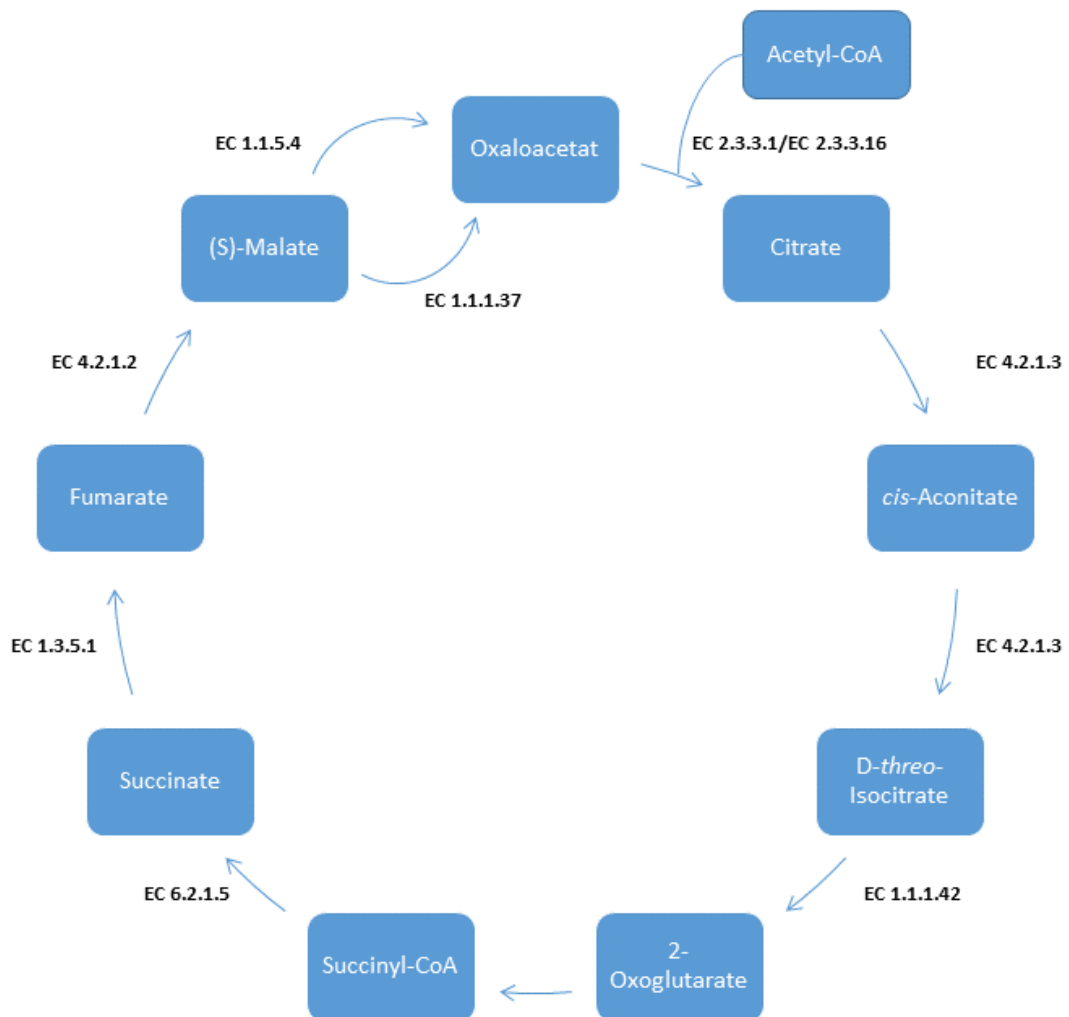


Abbildung 6: TCA Cycle I. Der obligatorische Vektor [2.3.3.16 - 2.3.3.1] - [4.2.1.3] - [1.1.1.42] - [6.2.1.5] - [1.3.5.1] - [4.2.1.2] beschreibt alle zwingend notwendigen Enzyme und direkte Alternativen einzelner Reaktionsschritte. Der Alternativvektor [1.1.5.4] --- [1.1.1.37] verweist auf Enzyme, bzw. im entsprechenden Fall auf Teilläste, die jeweils als Ersatz vorliegen können und mindestens einfach abgedeckt sein müssen.

Durch das erstellte Programm wird in einer weiterführenden Analyse der Organismengruppen explizit und definitiv unterschieden, welche Stoffwechselwege dem Metabolismus der einzelnen Vertreter zuteilwerden und welche nicht. Den Interpretationen der Verknüpfungen zu den Metadaten auf Enzymebene (Kapitel 3.5) wird so Unterstützung geboten und eine weiterführende Interpretation auf der komplexeren Ebene zusammenhängender Stoffwechselwege ermöglicht. (Kapitel 3.6)

3.3 Die Analysegruppen

In den folgenden Kapiteln werden zwei verschiedene Verfahren die enzymatischen Funktionen gewählter Organismengruppen mit den jeweiligen Metadaten verknüpfen und eine Grundlage für entsprechende Interpretationen schaffen. Die 334 untersuchten und analysierten Projektorganismen setzen sich zusammen aus 200 Archaeen und 134 Rhodobacteraceen. Diese beiden Gruppen werden im Folgenden getrennt auf Zusammenhänge der Enzym- und Pathwayannotationen mit den Metadaten-Kategorien untersucht. Dazu werden 132 Metadatensätze zu den Rhodobacteraceen recherchiert. (Kapitel 2.4) Für 164 verschiedene Vertreter der Domäne der Archaeen können ebenso Informationen für die Einordnung in die entsprechenden Kategorien zur Verfügung gestellt werden. (Kapitel 2.4, Anhang 1 und Anhang 2)

3.4 Die Enzym-Evolution der Archaeen

Die Enzym-Evolution der Archaeen wird nach dem Verfahren I untersucht. (Kapitel 2.9) Die Zuordnung von Enzymklassen zu allen archaealen Genen auf Grundlage der Homologie (BLAST, Kapitel 2.2) zu Sequenzen bekannter enzymatischer Funktionen aus BRENDA (Kapitel 1.9) dient im Folgenden der Interpretation von Zusammenhängen des Metabolismus dieser Organismen und ihrem Lebensraum, den dort vorherrschenden Bedingungen und evolvierter Entwicklungen.

Alle BLAST-Ergebnisse (Kapitel 2.9) über einem *E-Value*-Schwellenwert von kleiner als 0.001 und einer höheren Identität als 45% gehen in die Auswertung der metadatentypischen Enzyme der Archaeen mit ein. Die Enzyme, die auf diese Weise zu großen Anteilen nur spezifischen Metadaten, verknüpft mit den jeweiligen Organismen, zugeordnet werden können, sind in den Tabellen 6-11 aufgeführt. So werden im Folgenden individuelle Habitat-Relevanzen der Enzyme interpretiert.

Bakterien, Archaeen und Eukaryoten können unter extremsten Bedingungen einschließlich in sehr sauren, bzw. alkalischen Habitaten gefunden werden. [122] Der pH-Wert innerhalb einer Zelle allerdings ist generell im neutralen Bereich zu messen. [123] Wenn der pH-Wert des Habitats nun folglich einen Einfluss auf das Vorhandensein oder die Aktivität einer enzymatischen Funktion haben soll, gilt das vorrangig für extrazelluläre Enzyme. Andererseits gibt es auch Enzyme, die z.B. Komponenten der Zellwand synthetisieren. Diese können intrazellulär lokalisiert sein. Sie sind jedoch beteiligt an der Synthese von Komponenten der archaealen Zellmembran. Es ist dabei zu erwarten, dass diese eventuell besondere Strukturen aufweist, die diese Archaeen in Gegenwart der extremen pH-Werte überhaupt lebensfähig macht. [124]

Tabelle 6: Enzym-Evolution der Archaeen: **pH-Wert**. Die Tabelle zeigt auf Grund ihrer Häufigkeit typische Enzyme für entsprechend zugeordnete acidophile (Optimum pH 1-4), bzw. alkaliphile (Optimum von pH 9-14) Organismen. Der pH-Wertbereich entspricht dem Optimum-Durchschnitt aller Vertreter, die ein dem Enzym sequenzhomologes Gen aufweisen. Die Funktionen und Namen der Enzyme sind, wenn keine weitere Quelle angegeben ist, aus BRENDA übernommen.

saurer pH-Wertbereich: acidophil		
EC-Nummer	Bereich pH	Funktion/Name [20]
EC 6.3.4.18	1,9-2,6	5-(carboxyamino)imidazole ribonucleotide synthase, Schlüsselenzym der de novo Purin-Biosynthese[125]
EC 1.1.1.359	1,9-2,7	aldose 1-dehydrogenase [NAD(P)+]
EC 3.2.1.3	2-2,6	glucan 1,4- α -glucosidase
EC 3.4.23.42	2-2,8	thermopsin, Hydrolyse von Peptidbindungen [5]
EC 3.13.1.1	2,1-2,7	UDP-sulfoquinovose synthase
EC 1.2.1.75	2,2-2,9	malonyl-CoA reductase (malonate semialdehyde-forming)
EC 1.18.1.2		ferredoxin-NADP+ reductase
EC 4.2.1.140	2,4-3	gluconate/galactonate dehydratase
EC 1.16.1.1	2,6-3,3	in vivo Detoxifizierung von Quecksilber, bei hohen Temperaturen und Salzkonzentrationen [126]
EC 2.7.7.72	2,7-3,4	CCA tRNA nucleotidyltransferase
EC 3.2.1.20	2,9-3,5	α -glucosidase
EC 3.4.23.52		preflagellin peptidase
EC 2.6.1.97	3-3,6	archaeosine synthase
EC 3.2.1.21	3,3-3,9	β -glucosidase
EC 3.2.1.B26	3,5-4	<i>Sulfolobus solfataricus</i> β -glycosidase
EC 3.2.1.B33		<i>Sulfolobus shibatae</i> β -glycosidase
EC 3.2.1.B34		<i>Sulfolobus acidocaldarius</i> β -glycosidase
EC 4.2.1.9	3,8-4,2	dihydroxy-acid dehydratase
EC 4.2.1.120		4-hydroxybutanoyl-CoA dehydratase
EC 3.2.1.B28	3,8-4,3	<i>Pyrococcus furiosus</i> β -glycosidase
EC 5.3.4.1		Protein disulfide-isomerase
EC 4.2.1.150	4,1-4,5	short-chain-enoyl-CoA hydratase
alkalischer pH-Wertbereich: alkaliphil		
EC-Nummer	Bereich pH	Funktion
EC 1.1.1.179	9,5	D-xylose 1-dehydrogenase (NADP+)
EC 2.7.1.48		uridine kinase
EC 3.4.19.13		glutathione hydrolase
EC 4.1.3.36		1,4-dihydroxy-2-naphthoyl-CoA synthase
EC 4.2.1.55		3-Hydroxybutyryl-CoA dehydratase

Der Entner-Doudoroff-Weg

Einige Enzyme, die vor diesem Hintergrund überrepräsentativ in Organismen annotiert werden, die entweder mit einem pH-Optimum im stark sauren oder alkalischen Bereich verknüpft sind, listet Tabelle 6 auf. Eine Alternative zur Energiegewinnung durch den Abbau von Zuckern ist neben der Glycolyse bei vielen Bakterien und Archaeen der Entner-Doudoroff-Weg. [127] Die Archaeen entwickelten dabei drei verschiedene bekannte Modifikationen des Entner-Doudoroff-Weges. Auf der einen Seite gibt es die Variante, bei der die Entstehung von 2-Dehydro-3-Deoxy-6-Phosphogluconat als Edukt der Pyruvatabspaltung durch eine Aldolase ausbleibt. Hingegen entsteht 2-Dehydro-3-Deoxy-Gluconat, welches anschließend wiederum durch eine spezifische Aldolase zu Pyruvat und Glycerinaldehyd gespalten wird. Das entstandene

Aldehyd wird ferner durch die Katalyse entweder einer NAD(P)⁺- oder Ferredoxin-abhängigen Dehydrogenase zu Glycerat oxidiert. Folglich wird unter Verbrauch von einem ATP das Glycerat mit Hilfe einer Kinase zu 2-Phosphoglycerat, welches der Weiterverarbeitung zu Pyruvat durch glycolytische Enzyme zur Verfügung steht, umgesetzt. Dabei entsteht ein ATP. Diese Variante bezeichnet man auf Grund des Mangels an Netto-ATP-Gewinn und folglich keiner stattfindenden Phosphorylierung von ADP als nicht-phosphorylierenden Entner-Doudoroff-Weg. Er wird oft bei thermoacidophilen Archaeen beschrieben. [128] Eine zweite Variante, die zumeist bei halophilen Archaeen charakterisiert wird, unterscheidet sich durch eine Kinase-katalysierte Phosphorylierung des 2-Dehydro-3-Deoxy-Gluconat unter Verbrauch von einem ATP. Die darauffolgende Aldolspaltung zu Pyruvat und Glycerinaldehyd-3-phosphat ermöglicht in der anschließenden Glycolyse eine weitere Umsetzung zu Pyruvat mit der zweifachen Entstehung von ATP. Je nach Vorkommen der entsprechenden Enzyme beim jeweiligen Organismus entstehen letztendlich zwei Moleküle der Reduktionsäquivalente NADPH, NADH oder aber auch reduziertes Ferredoxin und netto ein ATP. Infolgedessen wird dieser Weg als halb-(semi-)phosphorylierend bezeichnet. [128,129] Archaeen, wie zum Beispiel die Sulfolobales, betreiben scheinbar gleichzeitig sowohl den nicht- und den semi-phosphorylierenden, genannt den verzweigten Entner-Doudoroff-Weg. Der wesentliche Unterschied besteht auch hier darin, dass kein ATP Gewinn generiert wird. Dies resultiert aus der direkten Oxidation von Glycerinaldehyd-3-phosphat zu 3-Phosphoglycerat unter der Verwendung einer nicht-phosphorylierenden Glycerinaldehyddehydrogenase. [128]

Allen drei Wegen ist anfangs der Katabolismus der Glucose, bzw. Galactose direkt zu einem Lacton und wiederum die Umsetzung des Gluconats oder Galactonats zum 2-Dehydro-3-Deoxy-Gluconat/2-Dehydro-3-Deoxy-Galactonat gemeinsam. Für diese Schritte stellt die Analyse nun die NAD(P)⁺ abhängige Aldose 1-Dehydrogenase (EC 1.1.1.359) und die Gluconat/Galactonat Dehydratase (EC 4.2.1.140) in Verbindung mit Organismen aus extrem sauren Umgebungen. Die Gluconat/Galactonat Dehydratase zum Beispiel hat eine breitere Substratspezifität als die Alternativen EC 4.2.1.6 und EC 4.2.1.39, die für Galactonat bzw. Gluconat spezifisch sind, indem sie beide Zucker als Substrat akzeptiert. [20] Da die Enzyme des Entner-Doudoroff-Weges allerdings alle intrazellulär agieren, ist der Zusammenhang nicht ganz eindeutig nachzuvollziehen. Eine Möglichkeit der Erklärung bietet zum Teil die evolutionäre Entwicklung der acidophilen Archaeen, indem diese expliziten Enzymfunktionen sich als Alternativen bei diesen Vertretern einfach behaupteten. [130–132]

Glycosidasen in Archaeen

Die vorläufig und neu klassifizierten Enzyme (Kapitel 1.5) der *Sulfolobus solfataricus* β -Glycosidase (EC 3.2.1.B26), *Sulfolobus shibatae* β -Glycosidase (EC 3.2.1.B33) und *Sulfolobus acidocaldarius* β -Glycosidase (EC 3.2.1.B34) zeichnen sich durch ihre breitere Spezifität im Gegensatz zur ursprünglichen β -Glucosidase (EC 3.2.1.21) aus. Neben Glucose können die Organismen aus dem sauren Milieu damit auch intrazellulär Oligomere mit z.B. Galactose, Fucose und Xylose hydrolysieren. [20] Die β -Glycosidase von *S. shibatae* spaltet im Unterschied zum entsprechendem Enzym aus *S. solfataricus* bevorzugt Laminaribiose [20], ein Abbauprodukt von Laminarin aus Braunalgen [133]. Die breitere Spezifität der Enzyme der Archaeen ist allerdings nicht nur auf extrem saure Habitate zu beziehen. Ferner ist dies archaeeentypisch. [134,135]

Proteasen in Archaeen

Weiterhin überrepräsentiert im sauren Milieu und mit breiter Substratspezifität, jedoch mit extrazellulärer Aktivität, können die Peptidasen Thermopsin (EC 3.4.23.42) und die *preflagellin peptidase* (EC 3.4.23.52) festgestellt werden. Sie spalten Signalpeptide zum Beispiel von Preflagellin zur Bildung von Flagellin, Pilin bis hin zu anderen zuckerbindenden Proteinen. [20] Sie sind unentbehrlich für die Biosynthese von Flagellin und Pilin in Sulfolobaceae und anderen acidothermophilen Archaeen. Wobei das Flagellin in Archaeen eher dem Typ-IV Pilin in Bakterien ähnelt als dem typischen Flagellin. [136] Peptidasen reagieren generell optimaler im sauren Bereich. [137] Das gilt ebenso für Eukaryoten und Bakterien. [20] Explizite Zelloberflächenstrukturen, wie z.B. Flagellen und Pili der Archaeen zeigen sich folglich auch in der Voraussetzung enzymatischer Funktionen in vorwiegend extrem sauren Habitaten. [136]

CO₂-Fixierung in Archaeen

Die Malonyl-CoA reductase (EC 1.2.1.75) und die 4-Hydroxybutanoyl-CoA Dehydratase (EC 4.2.1.120) gehören zum Pathway der Kohlenstoffdioxid-fixierung in Crenarchaeota. [20] Während der 3-Hydroxypropionat/4-Hydroxybutyrat Zyklus zur autotrophen Kohlenstofffixierung in *Metallosphaera sedula* (Sulfolobales, Crenarchaeota) und einigen Crenarchaeota nachgewiesen ist [138,139], konnte er für *S. solfataricus* bisher nur theoretisch postuliert werden. [140,141] Dieser Pathway ist sauerstofftolerant und demzufolge vorrangig bei aeroben Vertretern der Sulfolobales nachzuvollziehen. Aber auch anaerobe Spezies der Sulfolobales erfüllen alle enzymatischen Anforderungen. Ebenso kann gezeigt werden, dass ein unter anaeroben Bedingungen optimaler Dicarboxylate-4-Hydroxybutyrat Zyklus wie erwartet bei anaeroben Vertretern (*Desulfurococcales* und *Thermoproteales*), aber auch unter immerhin niedrigen Sauerstoffkonzentrationen bei den aeroben *Desulfurococcales* annotiert ist. [141]

Die Tatsache, dass diese Art der CO₂-Fixierung demzufolge nicht immer in den dafür optimalen Habitaten zu finden sind, wirft zwei Fragen auf. Handelt es sich hierbei um ein tiefer verwurzeltes „Überbleibsel“ der Evolutionsgeschichte dieser Spezies? [142] Oder sind diese enzymatischen Funktionen noch in einem anderen bisher ungeklärten Kontext denkbar? Gezeigt werden kann, dass die beiden genannten Schlüsselenzyme zumindest mit acidophilen Organismen verknüpft sind. Die evolutionsgestützte Theorie kann durch die Feststellung untermauert werden, dass ursprünglich viele *Sulfolobales* Spezies (*Sulfolobus solfataricus*, *Sulfolobus acidocaldarius*) als autotroph oder mixotroph beschrieben wurden. [143] Die Fähigkeit zur Kohlenstofffixierung ist nur bei Laborstämmen mit der Zeit während der fortwährenden Kultivierung auf nährstoffreichen Medien verloren gegangen. [141,144] Weiterhin scheinen sich verschiedene, jedoch alternative, Enzymfunktionen bei Archaeen aus sauren und alkalischen Habitaten entwickelt zu haben. So wird bei Organismen mit einem pH-Optimum im sauren Bereich von durchschnittlich etwa 4 der Schritt der Dehydratisierung von (S)-3-Hydroxybutyryl-CoA zu Crotonyl-CoA durch das Enzym EC 4.2.1.150 katalysiert. Die Alternative in Form der 3-Hydroxybutyryl-CoA Dehydratase (EC 4.2.1.55) ist bei Vertretern mit einem durchschnittlichen pH-Optimum von 9,5 zu finden.

Ferredoxin Reduktasen in Archaeen

Die Ferredoxin-NADP⁺ Reduktase (EC 1.18.1.2) wird vorrangig bei Organismen aus einem stark sauren Lebensraum annotiert. Entscheidende intrazelluläre Rollen können dem Recycling von Ferredoxin oder auch Flavodoxin bei Archaeen und Bakterien zugeordnet werden. Dieses Enzym ist also typisch für Organismen, die Photosynthese betreiben [20], kann jedoch auch Reaktionen der Stickstofffixierung [145] und im Steroidmetabolismus katalysieren. Sowohl die metabolische Nutzung von Stickstoff, als auch der Aufbau von spezifischen Lipiden für den Einbau in die Membran können somit mit acidophilen Habitaten in Verbindung gebracht werden. [130]

Die Quecksilberentgiftung in Archaeen

Unter stark sauren und heißen Umweltbedingungen kann die Löslichkeit von natürlich vorkommenden Mineralen beeinflusst werden. [146] Dementsprechend kann das sonst in Wasser praktisch unlösliche Quecksilbersulfid [147] zum Beispiel bei niedrigem pH-Werten leicht zur toxischen Form der Quecksilber(II)-Ionen (Hg²⁺) oxidiert werden. Dies erfordert demzufolge die Anpassung der dort lebenden Organismen [148] und in diesem Fall der Archaeen aus dem entsprechenden Habitat. Die *mercury(II) reductase* (EC 1.16.1.1) ermöglicht einen Resistenzmechanismus durch die Reduktion des Quecksilberions Hg²⁺ in die nicht oxidierte und weniger toxische Form Hg⁰. [126,148] Bei dieser zwei Elektronenreduktion werden NADPH als Elektronendonator und Flavin Adenin Dinucleotid (FAD) als Cofaktor verwendet. [149] Dabei kann die genetische Grundlage der Quecksilberreduktase auch in einer domänenspezifischen Form für die Archaeen nachgewiesen werden. [150,151]

DNA- und RNA-Modifikationen der Archaeen

Ein Beispiel für die spezifische Evolutionsentwicklung der Archaeen und in diesem Fall der acidophilen Vertreter zeigt die CCA tRNA Nucleotidyltransferase (EC 2.7.7.72). [152] Mit Hilfe des Enzymes werden an das 3'-Ende der finalen tRNA die drei Nukleinbasen Cytosin-Cytosin und -Adenin angehängen. Dies ist letztendlich von essentieller Bedeutung für die Translation der Transfer-RNA, da an dieser Stelle die entsprechende Aminosäure verestert und damit an die tRNA gebunden wird. [153] Diese Polymerasen sind allerdings in der Evolution zweimal unabhängig voneinander entstanden. So können genetisch zwei Klassen, einmal eine rein archaeentypische und jene der Bakterien und Eukaryoten, differenziert werden. Die Funktion ist hierbei jeweils deckungsgleich. Starke Unterschiede und somit der Hinweis auf die unabhängige Evolution zeigen die jeweiligen Sekundärstrukturen der einzelnen Domänen. [154,155]

Archaeosin ist ein nur bei Archaeen bekanntes und strukturell von Guanosin modifiziertes Nucleosid, welches wesentlich die Stabilität der tRNA-Struktur unterstützt. Dabei ist der Einbau an Position 15 der tRNA entscheidend [156,157], um mittels Wasserstoffbrückenbindungen und elektrostatischen Interaktionen der geladenen Seitenkette die Tertiärstruktur der TransferRNA zu stabilisieren. [158,159] Die Archaeosin Synthase (EC 2.6.1.97) zur Biosynthese des Archeosins ist den Daten zufolge nicht nur wie bisher publiziert sehr oft bei thermophilen Archaeen zu finden [160], sondern ist auch vorrangig gruppentypisch für die acidophilen Vertreter.

Der Abbau von Polysacchariden bei Archaeen

Die Verstoffwechslung komplexer Polysaccharide wie Cellulose [161] und Xylose [162] ist bei Archaeen zu großen Teilen aufgeklärt. So kann die β -Glucosidase (EC 3.2.1.21) mit archaeentypischer breiter Spezifität auch unter extremsten Bedingungen, niedrige pH-Werte inbegriffen, die Hydrolyse von Cellobiose zu zwei Molekülen Glucose katalysieren [20,163]. Ebenso können vorläufig charakterisierte Enzyme wie die *Pyrococcus furiosus* β -Glycosidase (EC 3.2.1.B28) für saure Habitate typisiert werden. Dieses Enzym produziert Mannose und Glucose hauptsächlich ausgehend von Mannooligosacchariden und Cellobiose. [20] Die Aufspaltung von Polysacchariden in ihre Einfachzucker als Energie- und Kohlenstofflieferant hat demzufolge mehrfach enzymatische Funktionen hervorgebracht, die explizit Archaeen in sauren Habitaten anwenden. Zugleich sind aber auch Enzyme wie die *D-xylose 1-dehydrogenase* (NADP^+) (EC 1.1.1.179) bei Vertretern nur aus stark alkalischen Bereichen gehäuft zu finden. Auch dieses Enzym unterstützt einen ersten Schritt des oxidativen Xyloseabbaus unter NADP^+ -Verbrauch. [21] Dies zeigt, dass sich die Verwendung solcher Kohlenstoffquellen bei den Archaeen sowohl in sauren und stark alkalischen Habitaten evolvieren konnte, sich jedoch verschiedene enzymatische Ausstattungen durchsetzten.

Weitere Auffälligkeiten der acidophilen Archaeen

Weiterhin ist bei einigen Enzymen, wie der Protein Disulfidisomerase (EC 5.3.4.1), womöglich zur Stabilisierung von Disulfidbindungen und der Proteinfaltung [20], die Verknüpfung zum acidophilen Habitat nicht ganz eindeutig zu klären und sollte in weiterführenden Untersuchungen verifiziert werden. Des Weiteren zählt dazu die α -Glucosidase (EC 3.2.1.20) die eine reaktionskinetisch sehr schnelle Hydrolyse von Oligosacchariden katalysiert [20]. Auch die *dihydroxy-acid dehydratase* (EC 4.2.1.9) mit einer Schlüsselfunktion beim Aufbau verzweigtkettiger Aminosäuren [20] obliegt bisher keiner klaren Notwendigkeit bei acidophilen Archaeen. Diese Tatsachen weisen entweder darauf hin, dass einige enzymatische Funktionen noch nicht gänzlich in Kontext mit Stoffwechselwegen speziell unter sauren Bedingungen gesetzt werden können, bzw. unterstreichen die evolutionäre Entwicklung der acidophilen Archaeen als Gruppe. [38,132,152,164]

Charakteristische Enzyme für alkaliphile Archaeen

Die Archaeen aus vorwiegend alkalischen Habitaten zeigen ebenso Anpassungen, vor allem, um die verfügbaren Kohlenstoffquellen zu nutzen und mit verwerteten Bausteinen möglichst effizient zu haushalten. So zum Beispiel weist die vorrangig bei diesen Vertretern annotierte Uridinkinase (EC 2.7.1.48) als eines der Schlüsselenzyme für den *Pyrimidine Ribonucleosides Salvage I* Stoffwechselweg [21] auf eine Möglichkeit hin, freie Purine aus Purinnukleotiden zu regenerieren und so Energie und Ressourcen zu sparen. Dieser Weg über das angesprochene Recycling ist gegenüber der *de novo* Biosynthese von Purinmononukleotiden erheblich energieeffizienter. [165,166]

Jedoch treten auch hier Enzyme ohne bisher eindeutigen Bezug zur alkaliphilen Lebensweise auf. Oft ist die katalytische Funktion bislang auch nur explizit bei Bakterien nachgewiesen. Ein Beispiel hierfür ist die 1,4-Dihydroxy-2-Naphthoyl-CoA Synthase (EC 4.1.3.36). Als Schlüsselenzym der Vitamin K Biosynthese ist es verantwortlich für die Überführung des o-Succinylbenzoyl-CoA zum 1,4-Dihydroxy-2-Naphthoyl-CoA. [167] Ebenso kann die enzymatische Hydrolyse von Glutathion (*glutathione hydrolase*, EC 3.4.19.13) bisher nur eindeutig bei Eukaryoten und Bakterien in einem Kontext mit der Aufnahme von Glutamin und Glutathion als Glutamatquelle gebracht werden. Die Aufnahme von extrazellulärem Glutamin, bzw. Glutathion als Stickstoffquelle ist direkt ohne vorherige Hydrolyse außerhalb der Zellen nicht möglich. [20,168]

Die Differenzierung acidophiler und alkaliphiler Archaeen offenbart zusammengefasst einige interessante Verknüpfungen. Besonders hervorzuheben ist hier eine alternative Reaktionskette für den Abbau von Glucose über ein Lacton im Entner-Doudoroff-Weg typisch für Archaeen in stark sauren Umgebungen. Ebenso typisch für diese Gruppe und im Zusammenhang mit den Habitatbedingungen zu sehen, ist ein spezifischer Entgiftungsweg von toxischem Quecksilber. Archaeen mit einem alkalischen pH-Optimum verfügen hingegen über sogenannte Salvage-Stoffwechselwege, die dem Recycling

aufwendig herzustellender Substrate dienen. Unterschiedliche enzymatische Funktionen und demzufolge alternativ entwickelte Wege in beiden Gruppen zeigen sich am Beispiel der CO₂-Fixierung.

Archaeen unter extremen Temperaturen

Mehr als dreiviertel der Erdoberfläche sind mit Ökosystemen im kalten Temperaturbereich von etwa 5°C belegt. Dazu zählen die Tiefen der Ozeane ebenso wie polare und alpine Regionen. Diese Habitate sind nicht unbelebt. Diese dauerhaft kalten Lebensräume sind erfolgreich von extremophilen Mikroorganismen besiedelt, zum Beispiel den psychophilen Archaeen. [169] Diese Eigenschaft, in solchen Umgebungen nahe dem Gefrierpunkt nicht nur zu überleben, sondern auch erfolgreiche Populationen auszubauen, benötigt eine Reihe an Anpassungen letztendlich, um den Metabolismus, so wie er von der Zelle benötigt wird, aufrecht zu halten. Dies erfordert sowohl Anpassungen der intrazellulären Proteine und Enzyme, als auch einen besonderen Aufbau der Zellmembranen. Der Stoffaustausch zwischen dem Zellinneren und der extremen Umgebung wird zum Beispiel durch den Einbau ungesättigter Fettsäuren mit kurzen Seitenketten aufrechterhalten, obwohl eine Barriere für molekulare Interaktionen nach Außen besteht. [170,171] Dieser Austausch muss genauestens angepasst sein, um einerseits genügend Metabolite und Energielieferanten bereitzustellen und andererseits das Temperaturoptimum zum Wachstum aufrecht zu halten. [172,173] Diese Organismen stellen sich zusammengefasst zwei Grundherausforderungen, die sie durch ihre Anpassungen bewältigen: Die niedrige Temperatur erfordert für diesen Bereich optimierte Enzyme, die entgegen der sonst typischen extrem niedrigen Aktivität unter 20°C bei den meisten Proteinen [174] dennoch durch eine erhöhte Konformationsflexibilität der Struktur [175] und eine herabgesetzte Energieschwelle zwischen der Aktivierung und der Katalyse [174] in diesen Temperaturbereichen arbeiten. Des Weiteren erfordert die temperaturbedingte Herabsetzung der Viskosität des umgebenden Wassers Adaptionen bezüglich der Stoffaufnahme und des Transports in die Zelle. [176]

Aber auch vor extrem hitzeintensiven Habitaten machen Archaeen nicht Halt. So leben einige thermophile Vertreter in hydrothermalen, bzw. sogenannten heißen Quellen unter Temperaturen von über 70°C bis hin zu über 100°C. [177] Auch dieses Umfeld erfordert Anpassungen, die letztendlich durch die breitgefächerte Vielfalt spezialisierter Enzyme und deren Struktur bewerkstelligt werden können. Typischerweise treten hier zum Beispiel gehäuft Disulfidbindungen und ionische Interaktionen in Erscheinung, um die Proteinstruktur trotz der hohen Temperatur zu erhalten. [178] Auch sind sehr oft stark hydrophobe Anteile vorhanden, die nach Außen exponiert den hydrophilen Kern schützen, um auch hier wieder der Struktur und Konformation Stabilität zu verleihen. [178,179]

3 Ergebnisse und Diskussion

Die Tabelle 7 listet die Enzyme auf, die bei Organismen der Archaeen-Domäne gefunden werden können, die wiederum im Durchschnitt ein Temperaturoptimum im eher psychrophilen (<41°C), bzw. thermophilen (90°C-100°C) Bereich aufweisen.

*Tabelle 7: Enzym-Evolution der Archaeen: **Temperatur**. Die Tabelle zeigt auf Grund ihrer Häufigkeit typische Enzyme für entsprechend zugeordnete psychrotolerante (Optimum unter 41°C), bzw. thermophile (Optimum von 90°C->100°C) Organismen. Der Temperaturbereich entspricht dem Optimum-Durchschnitt aller Vertreter, die ein dem Enzym sequenzhomologes Gen aufweisen. Die Funktionen und Namen der Enzyme sind, wenn keine weitere Quelle angegeben ist, aus BRENDA übernommen.*

psychrotolerant (<41°C):		
EC-Nummer	Bereich °C	Funktion/Name [20]
EC 6.1.1.26	33-35	pyrrolysine-tRNA ^{Pyl} ligase
EC 2.1.1.250	33-36	Methanmetabolismus
EC 2.7.11.32		[pyruvate, phosphate dikinase] kinase
EC 2.3.1.31	34-36	homoserine O-acetyltransferase
EC 2.5.1.49		O-acetylhomoserine aminocarboxypropyltransferase
EC 2.1.1.249	34-37	dimethylamine-corrinoid protein Co-methyltransferase
EC 4.1.1.B11		siroheme decarboxylase
EC 2.1.1.90	35-36	methanol-corrinoid protein Co-methyltransferase
EC 1.6.5.6	35-37	p-benzoquinone reductase (NADPH)
EC 2.7.13.3	35-38	histidine kinase
EC 1.8.4.11	36-38	peptide-methionine (S)-S-oxide reductase
EC 1.12.98.3	37-39	Methanosarcina-phenazine hydrogenase
EC 3.1.3.31	39-41	nucleotidase
thermophil (90°C-100°C):		
EC-Nummer	Bereich °C	Funktion/Name [20]
EC 2.7.7.74	90-91	1L-myo-inositol 1-phosphate cytidyltransferase
EC 2.7.8.34		CDP-L-myo-inositol myo-inositolphosphotransferase
EC 2.1.1.176	90-92	16S rRNA (cytosine967-C5)-methyltransferase
EC 3.5.99.10		2-iminobutanoate/2-iminopropanoate deaminase
EC 2.5.1.104	91-93	N1-aminopropylagmatine synthase
EC 2.7.1.40		pyruvate kinase
EC 2.7.2.2	92-94	carbamate kinase
EC 5.1.1.13		aspartate racemase
EC 3.1.3.70	93	mannosyl-3-phosphoglycerate phosphatase
EC 3.5.1.1		Aspartat und Asparagin Metabolismus [5]
EC 3.5.1.33	93-95	N-acetylglucosamine deacetylase
EC 3.5.1.105		chitin disaccharide deacetylase
EC 2.5.1.127	93-96	caldopentamine synthase
EC 3.2.1.165	94-97	exo-1,4-beta-D-glucosaminidase
EC 1.8.1.8	95	protein-disulfide reductase
EC 2.5.1.108		2-(3-amino-3-carboxypropyl)histidine synthase
EC 2.1.1.B43	98	tRNA (cytosine49-C5)-methyltransferase
EC 2.3.1.182		(R)-citramalate synthase

Die Methanogenese bei Archaeen aus kalten Habitaten

Die Verstoffwechslung von abgestorbener Biomasse unter anaeroben Bedingungen, die sogenannte Methanogenese, nutzen einige spezialisierte methanogene Archaeen der Abteilung der Euarchaeota als Energiequelle. [180] Dabei setzen die Vertreter der Ordnungen *Methanopyrales*, *Methanobacteriales*, *Methanococcales*, *Methanomicrobiales*, *Methanosarcinales*, *Methanocellales* und *Methanocellales* meist Kohlenstoffdioxid (CO_2) und Wasserstoff (H_2) zu Methan (CH_4) um. [180–183] Die meisten Habitate dieser Organismen sind bedingt durch die Tiefe, in der sie vorkommen, sehr oft sehr kalt und einem enormen Druck ausgesetzt. [183] Dort kann dann zusammen mit Wasser Methanhydrat entstehen, welches in der Forschung als möglicher Energieträger der Zukunft gesehen wird, sollte der rentable kommerzielle Abbau auf den Meeresböden gelingen. [184,185]

Der Verknüpfung zu diesen kalten Umweltbedingungen geschuldet, kann die *methanol-corrinoid protein Co-methyltransferase* (EC 2.1.1.90) vorrangig bei psychrophilen Archaeen gefunden werden. Dieses Enzym transferiert Methylgruppen von Methanol an einen kobaltbindenden Cofaktor [186] bei einer Stoffwechselwegvariante der Methanogenese (MetaCyc: methanogenesis from methanol [21]). Eine ganz ähnliche Rolle im Sinne der Methyltransferase tragen sowohl die *trimethylamine-corrinoid protein Co-methyltransferase* (EC 2.1.1.250) als auch die *pyrrolysine-tRNA^{PyI} ligase* (EC 6.1.1.26). [20] Pyrrolysin ist eine modifizierte Aminosäure, die als aktives Zentrum einiger Methyltransferasen fungiert. Es entsteht nicht durch die posttranslationale Modifizierung von Lysin, sondern wird vielmehr als Ganzes mit einer spezifischen Ligase in Proteine eingebaut. Die demzufolge pyrrolysinhaltige Methyltransferase benutzt Methylamine als Substrat und ist direkt beteiligt an der Methanogenese. [187,188]

Ebenfalls in Verbindung mit der Methanogenese in ähnlichem Kontext ist die *dimethylamine-corrinoid protein Co-methyltransferase* (EC 2.1.1.249) zu bringen. Allerdings katalysieren sowohl dieses Enzym, als auch die Enzymnummer EC 4.1.1.B11 (*siroheme decarboxylase*) archaeentypische alternative Reaktionsschritte bei der Häme-Biosynthese. Häme fungiert bei vielen Enzymen und sensorischen oder regulatorischen Proteinen als prosthetische Gruppe. [189] Diese Tetrapyrrole stellen jedoch ebenso eine essentielle Komponente der Elektronentransportkette bei der aeroben und anaeroben Atmung dar [189] und werden hier in Verbindung mit psychrophilen Archaeen gebracht.

Weitere typische Enzyme kalter Habitate

Die Geochemie mit den Faktoren Druck und Temperatur [190] kann ebenso wie die Biochemie des Metabolismus der einzelnen Organismen, zum Beispiel während der Zellatmung, die Bildung von reaktiven Stoffspezies begünstigen [191]. So entstehen im speziellen Fall für Zellen und Erbgut schädliche Formen des Sauerstoffs (**reactive oxygen species, ROS**), wie zum Beispiel freie Radikale, Peroxide, Ozon oder negativ geladene Halogenverbindungen. Als Folge entwickelten alle Domänen Eukaryota, Bacteria und Archaea einerseits

Signalempfänger für diesen oxidativen Stress und ein Schutzsystem aus enzymatischen Radikalfängern und Antioxidantien. Andererseits existieren Schutz- und Reparaturmechanismen für die DNA. [192–194] Das hier für psychrophile Archaeen herausgestellte Enzym *peptide-methionine (S)-S-oxide reductase* (EC 1.8.4.11) besitzt eine wichtige Funktion bei der Reparatur von oxidationsgeschädigten Proteinen. [20,195]

Für kalte Habitate werden noch weitere verschiedene enzymatische Funktionen typisiert. So ist mit Hilfe der *O-acetylhomoserine aminocarboxypropyl-transferase* (EC 2.5.1.49) Methanethiol oder Dimethylsulfid als einzige Schwefelquelle ausreichend, um Methionin zu synthetisieren. [196] Weiterhin scheinen sich besonders energiesparende Elektronentransportsysteme bei den psychrophilen Archaeen etabliert zu haben. Dementsprechend wird bei den Spezies der methanogenen *Methanosarcina* mit dem Enzym *methanosarcina-phenazine hydrogenase* (EC 1.12.98.3) der Protonen-/Elektronenaustausch über die Zellmembran hinweg katalysiert, um einen Protonengradienten aufzubauen und letztendlich ATP zu generieren. [197]

Zusätzlich treten noch weitere Funktionen gehäuft auf, die entweder eher in anderen Domänen zu erwarten sind, oder deren Rolle im Kontext mit dem Habitat nicht eindeutig zu klären ist. Die *[pyruvate, phosphate dikinase] kinase* (EC 2.7.11.32) ist bisher experimentell nur bei den Pflanzen, im speziellen den Magnoliophyta, nachgewiesen. Eine direkte Anwendung für Archaeen ist nicht klar zu erkennen. Da die ausgemachten Sequenzen vorrangig bei den *Methanosarcina* und *Methanococcoides* einen sehr guten E-Value nahe Null und eine Identität von mehr als 55% aufweisen, kann zumindest eine ähnliche enzymatische Funktion nicht ausgeschlossen werden. Dieses bifunktionale Enzym kann die *pyruvate, phosphate dikinase* (EC 2.7.9.1) sowohl phosphorylieren als auch dephosphorylieren. [198,199] Dieses Enzym wiederum ist bei Archaeen beschrieben [20] und steht im Zusammenhang mit der Kohlenstofffixierung bei Prokaryoten [24]. Andere Enzyme weisen zwar eine für Archaeen typische breite Substratspezifität auf, wie die *Nucleotidase* (EC 3.1.3.31), oder besitzen vielfältige Signalfunktionen in der Osmo-, Stickstoff oder Atmungsregulation (*histidine kinase*, EC 2.7.13.2) [20], deren Kontext in niedrig temperierten Habitaten ist jedoch nicht schlüssig.

Typische Enzyme heißer Habitate

Gleichermaßen wie die Anpassungen an niedrige Temperaturen ermöglichen auf der anderen Seite einige Enzyme wie zuvor schon erwähnt das Überleben unter sehr hohen Temperaturen. Das bifunktionale Enzym *1L-myo-inositol 1-phosphate cytidyltransferase* (EC 2.7.7.74/EC 2.7.8.34) und die damit katalysierte Biosynthese von di-myo-Inositolphosphat werden in der Forschung schon als Marker für die Adaption an heiße marine Umweltbedingungen genutzt. [200] Die Anreicherung der kompatiblen Solute di-myo-Inositolphosphat und Diglycerolphosphat als Antwort auf Hitze- und Osmostress, bzw. die dazu benötigten Enzymfunktionen sind genetisch bei den meisten Organismen auf einem Gen fusioniert. [201,202] Sie können allerdings auch individuell auf

separierten und nicht bifunktionalen Genen vorliegen (zum Beispiel *Aeropyrum pernix*, *Thermotoga maritima*, *Hyperthermus butylicus*). [202]

Auch die *mannosylglycerate* Biosynthese, ein negativ geladenes Osmolyt, trägt dazu bei, Proteine und deren Struktur auch bei hohen Temperaturen stabil zu halten. [203] Mit Hilfe der *mannosyl-3-phosphoglycerate phosphatase* (EC 3.1.3.70) wird die Hydrolyse des Zwischenproduktes *mannosyl-3-phosphoglycerate* zum Thermostabilisator *alpha-mannosylglycerate* katalysiert.

Besonders die DNA und RNA bedürfen weiteren und besonderen Schutzes bei hohen Temperaturen. [204,205] Dazu synthetisieren viele Archaeen Polyamine wie zum Beispiel Spermidin, Thermin und Caldine, die mit ihrer langkettigen und verzweigten Struktur der Denaturierung entgegenwirken, aber auch, in die Membran eingebaut, die Thermostabilität der Zelle an sich erhöhen. [206,207] Die Enzyme *N1-aminopropylagmatine synthase* (EC 2.5.1.104) [208] und *caldopentamine synthase* (EC 2.5.1.127) [209] sind am Aufbau der entsprechenden Polyamine direkt beteiligt.

Neben der Verwendung von sehr einfachen Verbindungen als Kohlenstoffquelle haben die Archaeen auch Möglichkeiten entwickelt, um komplexe Polysaccharide zu metabolisieren. (Kapitel 3.4: Der Abbau von Polysacchariden bei Archaeen). Dementsprechend treten bei den thermophilen Vertretern als weiteres Beispiel gehäuft Enzyme auf, die mit dem Chitinabbau in Verbindung stehen. Abgestorbene Meereskrustentiere wie zum Beispiel Krebsartige und andere Garnelen können als Kohlenstoffquelle für Archaeen dienen. [161] In diesem Zusammenhang ist beispielsweise die *N-acetylglucosamine deacetylase* (EC 3.5.1.33), die Amidbindungen im Chitinabbauweg (MetaCyc: *chitin degradation I (archaea)* [21]) hydrolysiert [210], zu nennen. Im gleichen Kontext stehen die *chitin disaccharide deacetylase* (EC 3.5.1.105) und die *exo-1,4-beta-D-glucosaminidase* (EC 3.2.1.165), die allesamt ein Temperaturoptimum von über 70°C aufweisen. [20]

Auch verschiedene andere Enzyme versuchen Protein- und vor allem DNA- oder RNA-Strukturen durch Modifikationen, bzw. durch die Übertragung funktioneller Gruppen stabil zu erhalten. Demgemäß sind zum Beispiel Mechanismen wie die Methylierung von Nucleinbasen in allen Domänen entstanden. [211,212] Besonders thermophile Archaeen weisen mehr *cytosine methyltransferasen* (EC 2.1.1.176 und EC 2.1.1.B43) auf als andere Vertreter. [213,214] Die genaue Rolle der Methylierung bei der Translation ist allerdings noch Gegenstand aktueller Forschungen. [215]

Daneben existieren bei den thermophilen Archaeen offensichtlich verschiedene Abwehrmechanismen gegenüber reaktiven Komponenten (Inaktivierung reaktiver Imine, *2-iminobutanoate/2-iminopropanoate deaminase* (EC 3.5.99.10) [216]) und anderen teils toxischen Zwischenprodukten. Die *carbamate kinase* (EC 2.7.2.2) reguliert die Weiterverarbeitung und verhindert die Thermodenaturation von Carbamoyl Phosphat, einem sehr instabilen und möglicherweise auch toxischen Vorläufer von Pyrimidin und Arginin. [217]

Mit der *protein-disulfide reductase* (EC 1.8.1.8) können Reduzierungen von Verbindungen mit Disulfidbindungen katalysiert werden. Proteine mit

3 Ergebnisse und Diskussion

Disulfidbindungen weisen in der Regel eine höhere Stabilität auch unter hohen Temperaturen auf. [218,219] Bei anderen Beispielen wie der *pyruvate kinase* (EC 2.7.1.40), der *aspartate racemase* (EC 5.1.1.13) oder der *asparaginase* (EC 3.5.1.1) ist zwar eine gute Thermostabilität nachgewiesen [20], allerdings ist entweder die biochemische Rolle noch nicht eindeutig bestimmt oder der direkte Bezug zum thermophilen Charakter unklar.

Sauerstoffmetabolismusabhängige Enzyme

Die Tabelle 8 zeigt einerseits Enzyme, die häufig in Archaeen mit einem aeroben Metabolismus ausgemacht werden, und andererseits Enzymklassen, die offensichtlich wesentlich häufiger den Vertretern des anaeroben Stoffwechsels zur Verfügung stehen. Die unterschiedlich ausgerichteten Bereitstellungsformen von Energie setzen an sich schon unterschiedliche Stoffwechselwege voraus. [220,221] Dennoch sind auch Differenzen einiger anderer enzymatischer Funktionen zu beschreiben.

*Tabelle 8: Enzym-Evolution der Archaeen: **Sauerstoffmetabolismus**. Die Tabelle stellt auf Grund ihrer Häufigkeit typische Enzyme für Organismen mit aerobem und anaerobem Metabolismus gegenüber. Die Differenz entspricht den „(%-aerobere Vertreter) - (%-anaerobere Vertreter)“. Die Funktionen und Namen der Enzyme sind, wenn keine weitere Quelle angegeben ist, aus der BRENDA übernommen.*

vorrangig aerob:		
EC-Nummer	Differenzbereich	Funktion/Name [20]
EC 2.3.1.9	61	acetyl-CoA C-acetyltransferase
EC 1.1.1.47	59	glucose 1-dehydrogenase [NAD(P)+]
EC 1.1.1.119	54	glucose 1-dehydrogenase (NADP+)
EC 1.3.1.83	48	geranylgeranyl diphosphate reductase
EC 2.3.3.16		citrate synthase (unknown stereospecificity)
vorrangig anaerob:		
EC-Nummer	Differenzbereich	Funktion/Name [20]
EC 2.1.1.77	-68	protein-L-isoaspartate(D-aspartate) O-methyltransferase
EC 5.4.99.25	-64	tRNA pseudouridine55 synthase
EC 1.2.7.1	(-50)-(-60)	pyruvate synthase
EC 1.5.98.1		Methan Metabolismus
EC 1.5.98.2		Methan Metabolismus
EC 1.8.98.1		CoB-CoM heterodisulfide reductase
EC 2.1.1.98		diphthine synthase
EC 2.3.1.101		formylmethanofuran-tetrahydromethanopterin N-formyltransferase
EC 2.3.3.13		2-isopropylmalate synthase
EC 2.4.2.57		AMP phosphorylase
EC 2.5.1.41		phosphoglycerol geranylgeranyltransferase
EC 2.5.1.77		7,8-didemethyl-8-hydroxy-5-deazariboflavin synthase
EC 2.7.2.3		phosphoglycerate kinase
EC 2.7.8.28		2-phospho-L-lactate transferase
EC 3.3.1.1		adenosylhomocysteinase
EC 3.1.21.B1		endonuclease Q
EC 3.1.99.B1		flap endonuclease-1
EC 3.5.4.4		adenosine deaminase
EC 3.5.4.39		GTP cyclohydrolase IV
EC 3.5.4.41		5'-deoxyadenosine deaminase
EC 4.1.1.23		Orotidine-5'-phosphate decarboxylase
EC 4.2.1.36		homoaconitate hydratase
EC 4.2.1.114		methanogen homoaconitase
EC 4.2.1.122		tryptophan synthase (indole-salvaging)
EC 6.1.1.3		threonine-tRNA ligase
EC 6.1.1.15		proline-tRNA ligase
EC 6.3.2.6		phosphoribosylaminoimidazolesuccinocarboxamide synthase
EC 6.3.4.22		tRNAIle2-agmatinylcytidine synthase
EC 6.5.1.1		DNA ligase (ATP)

Die *glucose 1-dehydrogenase (NADP⁺)* (EC 1.1.1.119) kommt mit einer Differenz von 54% häufiger bei Archaeen mit aerobem Metabolismus vor. Hierbei handelt es sich um ein weiteres alternatives Enzym, welches im Entner-Doudoroff-Weg anfangs die Umsetzung der Glucose, bzw. Galactose direkt zu einem Lacton katalysiert (Kapitel 3.4: Der Entner-Doudoroff-Weg). [20] Vor dem selben Hintergrund ist die wahlweise NAD⁺- oder NADP⁺-abhängige *glucose 1-dehydrogenase [NAD(P)⁺]* (EC 1.1.1.47) zu sehen. Sie stellt eine weitere Alternative für den gleichen Reaktionsschritt dar [222] und ist zu knapp 60% häufiger bei aeroben Archaeen zu finden.

Weitere Enzyme, wie die acetyl-CoA C-acetyltransferase (EC 2.3.1.9) nehmen eine fundamentale Funktion in vielen Abbauwegen von verschiedenen Aminosäuren, Fettsäuren und anderen verzweigtkettigen Stoffverbindungen ein. Das Enzym katalysiert den Transfer der Acylgruppe [20] und ist mit über 60% häufiger in aeroben Vertretern zu finden. Ebenso typisch für einen aeroben Stoffwechsel mit etwa 50% höherer Häufigkeit ist die *citrate synthase (unknown stereospecificity)* (EC 2.3.3.16) mit der Katalyse des Eingangsschrittes des TCA-Zyklus. [223] Dieser Stoffkreislauf dient vorwiegend bei aeroben Zellen dem oxidativen Abbau organischer Stoffe zur Energiegewinnung. [224]

Der Methanmetabolismus

Die Archaeen mit anaerobem Metabolismus weisen sehr häufig Enzyme aus dem Methanmetabolismus, dem mikrobiellen, anaeroben Abbau von Biomasse zur Energiegewinnung unter Bildung von Methan [225], auf. So sind die *methylenetetrahydromethanopterin dehydrogenase* (EC 1.5.98.1) und die *5,10-methylenetetrahydromethanopterin reductase* (EC 1.5.98.2) zwischen 50% und 60% häufiger vertreten und katalysieren wichtige Redoxschritte der Methanogenese ausgehend von Kohlenstoffdioxid. Ebenso in diesem Zusammenhang ist die *GTP cyclohydrolase IV* (EC 3.5.4.39) zu erwähnen. Typisch für methanogene Archaeen ist es an der Bildung einer Methanopterin-Vorstufe beteiligt. [226–228] Methanopterin fungiert als Coenzym zur Übertragung von C1-Körpern, biochemisch ganz ähnlich zu verschiedenen Folaten (Folsäure). [229–231]

Andere Enzyme finden wieder, gekennzeichnet durch ein recht breites Spektrum, Anwendung in verschiedenen Stoffwechselwegen. Die methanogene *homoaconitase* (EC 4.2.1.114) dient einerseits der Coenzym B Biosynthese, einem wichtigen Elektronenträger bei der Methanogenese. [232] Andererseits katalysiert es aber auch zusammen mit der *homoisocitrate dehydrogenase* (EC 1.1.1.87) Decarboxylierungen im sogenannten *2-oxoacid elongation pathway*. Dieser Stoffwechselweg baut die Kohlenstoffketten von Vorläuferintermediaten von Leucin, Isoleucin und dem schon erwähnten Coenzym B letztendlich durch eine Hydroxysäure-Isomerisation bis zum Endprodukt weiter aus. [233] Ebenso ist eine Beteiligung bei der Biosynthese von Lysin über den *α -aminoadipate* Stoffwechselweg nachgewiesen. [232]

Die direkte Verknüpfung des sauerstofffreien Habitats zum Metabolismus der betroffenen Archaeen zeigt sich noch in einigen weiteren Enzymbeispielen, die

bis zu 60% häufiger vorkommen als in der gegenüberstehenden Gruppe der aeroben Organismen. Allerdings ist der Zusammenhang weniger eindeutig als zum Beispiel bei funktionellen Zusammenhängen mit der Methanogenese und ist in der Literatur noch nicht eingehend beschrieben.

Habitattypische Enzyme

Die Zuordnung der Archaeen zu ihren jeweiligen Habitaten kann diesen entsprechend typische Enzyme hervorbringen. (Tabelle 9) Allerdings treten für den terrestrischen und aquatischen Lebensraum keine Enzyme besonders häufig auf, während sie in den anderen Gruppen gar nicht oder nur wenig vertreten sind. Weiterhin führen die vulkanischen Habitate, die heißen und die Schwefelquellen durch zumeist hohe vorherrschende Temperaturen, unvermeidlich zu einer Überschneidung mit den schon erläuterten Ergebnissen der thermophilen Archaeen. Dennoch können auch hier weitere Enzyme mit ähnlichem Kontext typisiert werden.

*Tabelle 9: Enzym-Evolution der Archaeen: **Habitat**. Die Tabelle zeigt auf Grund ihrer Häufigkeit typische Enzyme für Organismen aus dem entsprechend zugeordneten Habitat. Der Prozentbereich entspricht dem Anteil gegenüber allen bekannten Vertretern des jeweiligen Habitats. Die Funktionen und Namen der Enzyme sind, wenn keine weitere Quelle angegeben ist, aus BRENDA übernommen.*

heiße Quellen:		
EC-Nummer	%-Bereich	Funktion/Name [20]
EC 3.1.3.11	80-89	fructose-bisphosphatase
EC 3.6.4.12		DNA helicase
EC 4.1.1.50		adenosylmethionine decarboxylase, Polyaminsynthese [234]
EC 6.3.4.2		CTP synthase (glutamine hydrolysing)
EC 6.3.5.7		glutaminyl-tRNA synthase (glutamine-hydrolysing)
vulkanisch:		
EC-Nummer	%-Bereich	Funktion/Name [20]
EC 3.1.3.11	80-89	fructose-bisphosphatase
EC 3.6.4.12		DNA helicase
EC 4.1.1.50		adenosylmethionine decarboxylase, Polyaminsynthese [234]
EC 6.3.5.7		glutaminyl-tRNA synthase (glutamine-hydrolysing)
Schwefelquellen:		
EC-Nummer	%-Bereich	Funktion/Name [20]
EC 2.4.2.28	90-100	S-methyl-5'-thioadenosine phosphorylase
EC 3.6.4.12		DNA helicase
EC 4.1.1.50	80-89	adenosylmethionine decarboxylase, Polyaminsynthese [234]
EC 1.4.1.4		glutamate dehydrogenase (NADP+)
EC 6.1.1.11		serine-tRNA ligase
Schmutzwasser:		
EC-Nummer	%-Bereich	Funktion/Name [20]
EC 1.12.98.1	90-100	coenzyme F420 hydrogenase
EC 2.1.1.86		tetrahydromethanopterin S-methyltransferase
EC 2.6.1.83		LL-diaminopimelate aminotransferase
EC 2.8.4.1		coenzyme-B sulfoethylthiotransferase
EC 3.5.4.27		methenyltetrahydromethanopterin cyclohydrolase
EC 4.1.1.23		Orotidine-5'-phosphate decarboxylase
EC 6.2.1.2		butyrate-CoA ligase
EC 1.2.99.9	80-89	formate dehydrogenase (coenzyme F420)

Die heißen Umgebungstemperaturen bringen neben vielen Schutzmaßnahmen auch die Notwendigkeit für Reparaturmechanismen des DNA- und RNA-Erbguts mit (Kapitel 3.4: Typische Enzyme heißer Habitate). Die *DNA helicase* (EC 3.6.4.12) ist neben dem Entwinden der DNA-Doppelstränge [235] auch in Zusammenhang mit einem solchen Mechanismus bei thermophilen Archaeen [236] und demzufolge auch mit den vulkanischen Habitaten und den heißen, bzw. Schwefelquellen zu bringen. Die Bildung von Polyaminen ebenfalls zur Stabilisierung von DNA, aber auch zum Einbau in eine hitzeresistente Membran, wird durch die katalytische Wirkung der *adenosylmethionine decarboxylase* (EC 4.1.1.50) unterstützt. Das bifunktionale Enzym *fructose-bisphosphatase* (EC 3.1.3.11/EC 4.1.2.13) katalysiert in der Gluconeogenese der meisten Archaeen zwei aufeinanderfolgende Reaktionsschritte, fällt durch eine hohe Thermostabilität auf [237] und unterliegt den Daten zufolge einer Verknüpfung zu den Bedingungen in heißen Quellen und vulkanischen Gebieten. Die Gluconeogenese dient dem zusätzlichen Aufbau von Glucose zumeist aus Pyruvat. [13]

In schwefelverbindungshaltigen, heißen Umgebungen treten neben einigen DNA-Reparaturenzymen (z.B. EC 3.6.4.12 und EC 6.1.1.11) auch Enzyme häufiger auf, die hier zur Reaktion auf Hitze und oxidativen Stress dienen. Für die *glutamate dehydrogenase (NADP⁺)* (EC 1.4.1.4) kann diese Funktion bei der Hefe *Saccharomyces cerevisiae* ebenfalls nachgewiesen werden. [238]

In Schmutz- und Abwasser lebende Archaeen haben einige enzymatische Funktionen zur Verfügung, die gehäuft nur bei diesen Vertretern genetisch codiert sind. Auch bei vielen dieser Vertreter wird bevorzugt Energie aus Stoffwechselwegvarianten der Methanogenese gewonnen. So treten Coenzyme wie zum Beispiel die *coenzyme F420 hydrogenase* (EC 1.12.98.1) [239] und verschiedene Transferasen wie die *tetrahydromethanopterin S-methyltransferase* (EC 2.1.1.86) [240] und die *coenzyme-B sulfoethylthiotransferase* (EC 2.8.4.1) [241] wesentlich häufiger auf als in anderen Habitaten. Der Zusammenhang anderer Enzyme kann hingegen nicht klar erfasst werden. So lässt das Enzym *butyrate-CoA ligase* (EC 6.2.1.2), welches in Eukaryoten tragende Rollen in der Konjugation von Glycin [242] einnimmt, bei dem aber auch die enzymatische Aktivität bei der Biosynthese von Antibiotika (Penicillin) nachgewiesen ist [243], nur Raum für Vermutungen im Kontext mit im Schmutzwasser lebenden Archaeen, die weiter untersucht werden müssen.

Die Halotoleranz-verknüpften Enzyme

Die Anpassung an verschiedene Salzkonzentrationen der jeweiligen Umgebung erfordert nicht nur morphologische Anpassungen, sondern ebenso sowohl sensorische als auch regulatorische Einrichtungen im Metabolismus der Organismen. Diesbezüglich werden auch enzymatische Funktionen, die unter den jeweiligen Bedingungen aktiv und stabil sind, benötigt. [244] (Tabelle 10)

*Tabelle 10: Enzym-Evolution der Archaeen: **Halotoleranz**. Die Tabelle zeigt auf Grund ihrer Häufigkeit typische Enzyme für Organismen mit der entsprechend zugeordneten Halotoleranzgruppe. Es wird unterschieden zwischen gering (1-6% NaCl-Konzentration), medium (6-15% NaCl-Konzentration) und extrem (>15% NaCl-Konzentration) halotolerant. Der Prozentbereich entspricht dem Anteil gegenüber allen bekannten Vertretern der jeweiligen Toleranzgruppe. Die Funktionen und Namen der Enzyme sind, wenn keine weitere Quelle angegeben ist, aus BRENDA übernommen.*

gering halotolerant:		
EC-Nummer	%-Bereich	Funktion/Name [20]
EC 3.1.99.B1	90-100	flap endonuclease-1
EC 6.1.1.15		proline-tRNA ligase
EC 6.5.1.1		DNA ligase (ATP)
EC 1.2.7.1		pyruvate synthase
EC 3.3.1.1		adenosylhomocysteinase
EC 5.4.99.25		tRNA pseudouridine55 synthase
medium halotolerant:		
EC-Nummer	%-Bereich	Funktion/Name [20]
EC 2.8.1.7	80-89	cysteine desulfurase
EC 3.1.3.31		nucleotidase
EC 5.4.99.B4		23S rRNA pseudouridine(helix 69) synthase
extrem halotolerant:		
EC-Nummer	%-Bereich	Funktion/Name [20]
EC 2.3.3.10	90-100	hydroxymethylglutaryl-CoA synthase
EC 2.8.1.7		cysteine desulfurase
EC 3.1.3.31		nucleotidase
EC 4.4.1.16		selenocysteine lyase
EC 5.4.99.B4		23S rRNA pseudouridine(helix 69) synthase
EC 6.3.2.2	80-89	Glutamate-cysteine ligase
EC 1.1.1.343		phosphogluconate dehydrogenase (NAD+-dependent, decarboxylating)
EC 1.8.1.4		dihydrolipoyl dehydrogenase
EC 1.11.1.21		catalase-peroxidase
EC 4.2.1.82		xylonate dehydratase

Schon Vertreter mit einer Toleranz gegenüber geringen Salzkonzentrationen (1-6% NaCl-Konzentration) weisen der Häufigkeit nach typische Enzyme auf. Auch hier etablierten sich demzufolge im Laufe der evolutiven Entwicklung verschiedene Faktoren und Regulatoren für DNA-Replikations- (*tRNA pseudouridine55 synthase* (EC 5.4.99.25) [245]) und Reparaturmechanismen (*flap endonuclease-1* (EC 3.1.99.B1) [246]). [47,48]

Die extrem halophilen Archaeen (>15% NaCl-Konzentration) besitzen mit der *cysteine desulfurase* (EC 2.8.1.7) und der *selenocysteine lyase* (EC 4.4.1.16) die besondere Möglichkeit, bestimmte Cofaktoren mit einem Eisen-Schwefel-Kern zu formieren [247], und dies sogar ohne direkten Zugang zu einer Schwefelquelle (S₀) [248]. L-Cysteine dient in diesem Fall als essentielle Schwefelquelle für die

sogenannten Eisen-Schwefel-Cluster der Coenzyme. [248] Diese wirken meist als Elektronentransferreaktanten oder zur Generierung aktiver Radikale. [249] Des Weiteren steht die *phosphogluconate dehydrogenase (NAD⁺-dependent, decarboxylating)* (EC 1.1.1.343) für eine für halotolerante Organismen typische Dehydrogenase im oxidativen Teil des Pentose-Phosphat-Stoffwechselweges zur Generierung von NADPH bei dem Wachstum auf Glucose als Hauptkohlenstoffquelle. [250]

Einige andere Enzymklassen der extrem halotoleranten Archaeen beschreiben der Funktion nach weitere regulierende oder schützende RNA- und DNA-verbundene Enzyme. Andere weisen jedoch auch auf Reaktionsschritte verschiedener Abbauewege, wie den Xylose Abbau, *xylonate dehydratase* (EC 4.2.1.82) hin. [162] Die meisten dieser angesprochenen enzymatischen Funktionen werden vor allem bei den extrem halotoleranten Spezies der *Haloferax* nachgewiesen.

Die Differenzierung des Metabolismus in chemotroph, lithotroph, organotroph, autotroph und heterotroph eignet sich nach diesem Verfahren weder untereinander noch gegenüber anderen Metadaten als Kriterium. Folglich lassen sich keine gruppentypischen Enzyme nennen.

Methanogene Archaeen

Die Produkte der Methanogenese dienen einem Großteil der Euarchaeota als Energiequelle. [183] Diverse typische Enzyme sind in diesem Kapitel schon im Zusammenhang mit anderen Metadaten genannt (Der Methanmetabolismus). Daher führt die Suche nach vorrangig bei methanogenen Archaeen annotierten Enzymen (Tabelle 11) zu vielen eben dieser schon erwähnten Beispiele oder sie sind im gleichen Kontext bezüglich verschiedener Cofaktoren zu sehen (EC 2.5.1.77, EC 2.7.8.28: Cofaktor F420).

*Tabelle 11: Enzym-Evolution der Archaeen: **methanogene Archaeen**. Die Tabelle zeigt auf Grund ihrer Häufigkeit typische Enzyme für methanogene Archaeen. Der Prozentbereich entspricht dem Anteil gegenüber allen bekannten Vertretern, bei denen die Methanogenese nachgewiesen ist. Die Funktionen und Namen der Enzyme sind, wenn keine weitere Quelle angegeben ist, aus BRENDA übernommen.*

methanogen:		
EC-Nummer	%-Bereich	Funktion/Name [20]
EC 1.5.98.1	95-100	Methanmetabolismus
EC 1.5.98.2		Methanmetabolismus
EC 2.3.3.13		2-isopropylmalate synthase
EC 2.5.1.77		7,8-didemethyl-8-hydroxy-5-deazariboflavin synthase
EC 4.2.1.36		homoaconitate hydratase
EC 2.7.8.28	90-95	2-phospho-L-lactate transferase
EC 4.1.1.23		Orotidine-5'-phosphate decarboxylase
EC 4.2.1.114		methanogen homoaconitase
EC 4.3.3.7		4-hydroxy-tetrahydrodipicolinate synthase
EC 6.1.1.27		O-phospho-L-serine-tRNA ligase
EC 3.1.21.B1	80-89	endonuclease Q
EC 3.5.4.27		methenyltetrahydromethanopterin cyclohydrolase
EC 4.1.99.12		3,4-dihydroxy-2-butanone-4-phosphate synthase
EC 6.1.1.21		histidine-tRNA ligase
EC 6.3.2.6		phosphoribosylaminoimidazolesuccinocarboxamide synthase

Dessen ungeachtet treten auch Enzyme mit anderen Zusammenhängen hier gehäuft in Erscheinung. Die *3,4-dihydroxy-2-butanone-4-phosphate synthase* (EC4.1.99.12) ist beteiligt an der Flavin Biosynthese. [251] Das gebildete Riboflavin ist Ausgangsstoff für wichtige Flavin-Cofaktoren wie FMN und FAD, welche in sehr vielen Redoxreaktionen in den Zellen als Elektronenüberträger dienen. [252,253]

Die Evolution domänentypischer Enzyme

Die Proteobakterien gliedern sich in die zwei Domänen der Archaeen und Bacteria. Beide Domänen teilen sich marine und nicht marine Habitate, während viele Vertreter der Archaeen auch in sehr extremen Bedingungen überleben können. Auf Grund der gemeinsamen Evolutionsgeschichte gibt es viele Gemeinsamkeiten [254,255], aber auch Unterschiede, die sich vor allem in unterschiedlich aufgebauten Zellwänden und in Zusammenhängen der Proteinbiosynthese zeigen. [38,256] Die daraus ebenfalls resultierenden Unterschiede im Spektrum der enzymatischen Funktionen sind im Folgenden untersucht.

Vierzehn Enzyme weisen über alle E-Value-Stufen hinweg im Median eine Differenz der Domänenanteile im Vergleich zu den Bakterien von größer als 40 auf. (Tabelle 12, linke Spalte) Diese enzymatischen Funktionen haben folglich einen archaealen Ursprung, bzw. sind sie im Spezialfall der Domäne der Archaeen eigen. Die Abbildung 7 stellt ein solches Beispiel anhand der Chaperonin ATPase (EC 3.6.4.9) dar. Über alle E-Value-Stufen von 10^{-3} bis kleiner als 10^{-200} hinweg werden etwa in 60% der Archaeen-Domäne Sequenzen mit der entsprechenden Homologie gefunden. Bei den Bakterien wird ein Maximalwert von knapp unter 20% nicht überschritten. Bei kleineren E-Values und somit auch größerer Identität und Sequenzhomologie können nur noch 10% der Bakterien mit diesem Enzym annotiert werden. Chaperone werden im Allgemeinen dazu benötigt, aufgebauten Aminosäureketten, bzw. Proteinen zu ihrer physiologischen Sekundärstruktur zu verhelfen. Die Unterstützung bei der Faltung von Proteinen an sich ist nicht archaeentypisch und lässt sich in allen Domänen belegen. [257] Gerade in den extremen Lebensräumen der Archaeen werden sowohl zum Aufbau und Schutz vieler Proteine, als auch zum Erhalt deren Struktur beispielsweise entsprechende hitzebeständige Chaperontypen benötigt. Weiterhin kann demzufolge die Proteinfaltung trotz oxidativem Stress, unter stark sauren oder alkalischen Bedingungen, aber auch in Anwesenheit zell- und proteinschädigender Substanzen überhaupt ermöglicht werden. Aus diesem Grund können in den Archaeen domäneneigene Chaperone nachgewiesen werden. [258] Das Beispiel der Chaperonin ATPase zeigt infolgedessen die spezielle Anpassung der entsprechenden Enzyme thermophiler Archaeen. Der hier angesprochene Typ II mit erhöhter Hitzeresistenz kommt vorrangig in Archaeen und Eukaryoten vor. Ein Pendant mit entsprechender Sequenzähnlichkeit können auch Bakterien aufweisen (Typ I). [259] In beiden Fällen werden ATP-abhängige Proteinfaltungen unterstützt, während der Archaeentyp strukturelle Besonderheiten aufweist. [20,260]

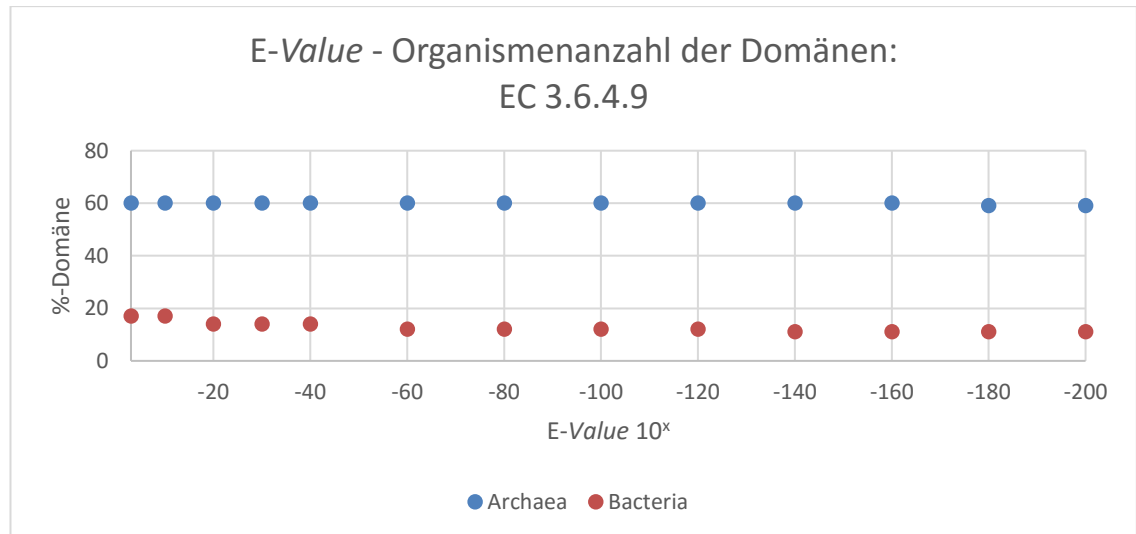


Abbildung 7: Archaeentypisches Enzym auf Grund der Domänenanteile zugeordneter BLAST-Treffer über alle E-Value-Stufen.

Weitere archaeentypische Beispiele sind in Tabelle 12 aufgelistet. Darunter sind Enzyme wie die *sn-glycerol-1-phosphate dehydrogenase* (EC 1.1.1.261), die ausschließlich bei Archaeen beschrieben sind. Dieses Enzym katalysiert Reaktionen zum Aufbau strukturtypischer Phospholipide der Zellmembran der Archaeen. [261] Ebenso ist die *tRNA-guanine15 transglycosylase* (EC 2.4.2.48) an der Archaeosin-Biosynthese beteiligt und damit einzigartig für diese Domäne (Kapitel 3.4: DNA- und RNA-Modifikationen der Archaeen). [262] Viele Enzyme sind allerdings im Kontext mit der Proteinfaltung und Strukturhaltung von DNA und RNA zu sehen. Dabei sind einige für Archaeen spezifisch und dazu fähig, unter den entsprechenden harschen Bedingungen zu agieren, so zum Beispiel die *chaperonin ATPase (protein-folding, protecting from aggregation, protein stabilizing)* (EC 3.6.4.B10). Diese schützt andere Enzyme nachgewiesener Maßen vor thermaler Aggregation [263] oder erhält die Funktion halophiler Proteine unter nicht optimal niedrigen Salzkonzentrationen [264]. Die *RadA recombinase* (EC 3.6.4.B7) und der *clamp-loader complex* (EC 3.6.4.B8) unterstützen und fördern die korrekte DNA-Rekombination trotz Temperaturunten nahe dem Schmelzpunkt von DNA. [265,266] Dahingegen sind andere Reparaturmechanismen offensichtlich zu hohen Anteilen bei den Archaeen annotiert, sie haben sich allerdings auch zumindest teilweise bei den Bakterien und Eukaryota evolviert. So repariert eine *DNA ligase (ATP)* (EC 6.5.1.1) Brüche des DNA-Doppelstranges und erhält somit die Integrität. [267,268] Dieser Mechanismus ist ATP abhängig. Nur bei Archaeen ist ein enzymatisches Pendant (*DNA ligase (ATP or NAD⁺)*) (EC 6.5.1.6) bekannt, welches eine breitere Substratspezifität aufweist und auch oxidierte Nicotinamid-Adenin-Dinukleotide (NAD⁺) verwendet. [269,270]

3 Ergebnisse und Diskussion

Demgegenüber werden dreizehn bakterientypische Enzyme mit einer im Median um 50 größeren Differenz des prozentualen Domänenanteils bei den Bakterien erkannt. (Tabelle 12, rechte Spalte) Dabei werden allerdings zumindest in Anbetracht geringerer Sequenzähnlichkeit (größerer E-Value) auch Sequenzen von Vertretern der Archaeen ebenfalls hervorgebracht. Am Beispiel der 3-Isopropylmalat Dehydrogenase (EC 1.1.1.85) in Abbildung 8 ist zu erkennen, dass beim BLAST (Kapitel 1.9) mit allen bereits mit dem Enzym annotierten Sequenzen aus der BRENDA-Datenbank (Kapitel 1.9) im E-Valuebereich von 10^{-3} bis 10^{-120} etwa 80% aller bakteriellen Organismen ein Gen mit hoher Sequenzähnlichkeit aufweisen. Hingegen werden nur bei unter 20% der Archaeen Vertreter gefunden. In kleineren E-Valuebereichen ($\leq 10^{-140}$) können immer noch bei ca. 40% der Bakterien Übereinstimmungen erkannt werden, während keine Vertreter der Archaeen mehr zur Ergebnismenge gehören. Dieses Enzym (EC 1.1.1.85) katalysiert die oxidative Decarboxylierung von 2-Hydroxycarbonsäuren und spielt eine Schlüsselrolle bei der Leucin und Isoleucin Biosynthese [20]. Die Verwendung verzweigter Aminosäuren zur weiteren Biosynthese verzweigter Isopreneinheiten zum Einbau in die Zellmembran (Mevalonat-Stoffwechselweg) ist bei vielen Bakterien bekannt. Da diese Strukturen bei den extremophilen Archaeen so immens wichtig für die Stabilität der Zellmembran sind, wird vermutet, dass einige Alternativen als Grundlage für mehrere höchst effiziente Biosynthesewege existieren. Der Aufbau von Leucin als Kohlenstoffquelle für die Synthese solcher Isoprenoid-basierter Fette kann bisher bei einigen halophilen Vertretern nachgewiesen werden. [271] Dies bekräftigt, dass nicht alle Archaeen Leucin als Kohlenstoffquelle verwenden müssen, es jedoch einige Vertreter gibt (20% der Domäne), die diese Alternative nutzen und somit auf die 3-Isopropylmalat Dehydrogenase als Schlüsselenzym angewiesen sind.

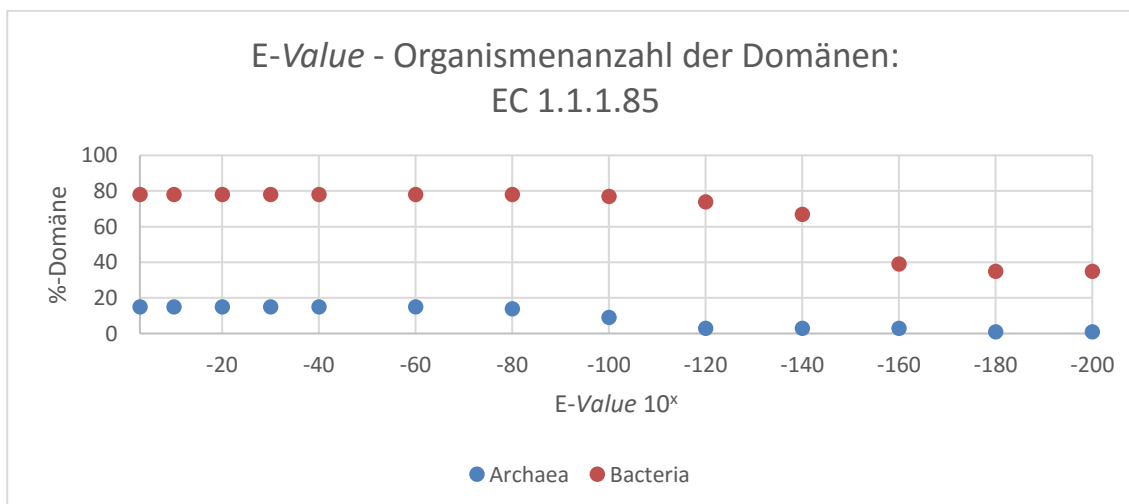


Abbildung 8 Bakterientypisches Enzym auf Grund der Domänenanteile zugeordneter BLAST-Treffer über alle E-Value-Stufen.

Unter den bakterientypischen Enzymen aus Tabelle 12 ist weiterhin noch eine *histidine kinase* (2.7.13.3) hervorzuheben. Dieses Enzym ist in die Signalweiterleitung der Chemotaxis bei Bakterien involviert. [272–274] Weiterhin können bakterienspezifische Transportsysteme herausgestellt werden. Die

bacterial protein-transporting ATPase (EC 3.6.3.B1) katalysiert den sekretorischen Transport über die Membrangrenzen der Zelle hinaus. [275] Ergänzend sind einige Enzyme wie die Protease *Endopeptidase Clp* (EC 3.4.21.92) aufgelistet, die ebenfalls ausschließlich unter den Bakterien und Eukaryoten weit verbreitet sind und eine geringe Substratspezifität aufweisen. [20,276]

Auch einige andere Enzymfunktionen sind allerdings zumindest teilweise bei den Archaeen vertreten (z.B. *methylenetetrahydrofolate reductase [NAD(P)H]* (EC 1.5.1.20), *dihydroorotase* (EC 3.5.2.3)). Dies unterstreicht im gewissen Rahmen die gemeinsame Evolutionsgeschichte und den eventuell möglichen horizontalen Gentransfer im Laufe der Zeit. [277,278] Jedoch können auch Konvergenzen zu Tage treten. Der Reparatur gebrochener DNA-Stränge kann auch bei den Bakterien mit der *DNA ligase (NAD⁺)* (EC 6.5.1.2) Folge geleistet werden. Auch hier ist wie beim zuvor erläuterten archaeentypischen Gegenüber NAD⁺ als Cofaktor einsetzbar. [279] Dieses Enzym kann allerdings mit den vorliegenden Daten nur bei Bakterien annotiert werden.

Tabelle 12: Domänentypische Enzyme. Als domänentypische Enzyme wurden jene klassifiziert, die über alle E-Value-Stufen hinweg im Median den größeren Prozentanteil in der jeweiligen Domäne einnehmen und in der Differenz über dem Schwellenwert von 40 (Archaea), bzw. 50 (Bacteria) liegen. Enzyme in „fett“ können noch keinem MetaCyc-Pathway mit vollständig klassifizierten Enzymen zugeordnet werden.

archaeentypische Enzyme		bakterientypische Enzyme	
EC-Nummer	Median(Differenz)	EC-Nummer	Median(Differenz)
EC 1.1.1.261	40	EC 1.1.1.85	-63
EC 2.4.2.48	50	EC 1.2.1.11	-58
EC 2.7.7.6	41	EC 1.5.1.20	-68
EC 3.6.1.3	41	EC 2.5.1.19	-69
EC 3.6.4.8	41	EC 2.5.1.55	-56
EC 3.6.4.9	48	EC 2.7.13.3	-89
EC 3.6.4.B10	55	EC 3.4.21.92	-68
EC 3.6.4.B7	58	EC 3.5.2.3	-50
EC 3.6.4.B8	53	EC 3.6.3.B1	-58
EC 3.6.5.4	44	EC 4.2.1.11	-64
EC 3.6.5.6	41	EC 4.3.3.7	-52
EC 6.1.1.23	53	EC 5.1.1.1	-76
EC 6.5.1.1	50	EC 6.5.1.2	-50
EC 6.5.1.6	50		

Einem Großteil der in diesem Abschnitt angesprochenen Enzyme kann in MetaCyc kein Stoffwechselweg mit allen vollständig klassifizierten enzymatischen Funktionen zugeordnet werden. Ein Grund dafür kann demzufolge die fehlende Klassifizierung eines beteiligten Enzyms sein. Oft ist es jedoch der Tatsache geschuldet, dass sehr viele dieser Enzyme an regulatorischen Prozessen und der Biosynthese der DNA und RNA beteiligt sind. Diese sind ebenso wie enzymatische Reparaturmechanismen keinem Stoffwechselweg des Metabolismus zugeordnet. Dennoch sind sie unverzichtbar für die Funktionalität der Zelle und somit letztendlich der Lebensfähigkeit der Organismen.

Zusammenfassung

Alles in Allem lassen sich durch das Verfahren I (Kapitel 2.9), wie beschrieben, deutliche Verknüpfungen von enzymatischen Funktionen zum pH-Wert- und Temperaturoptimum der jeweiligen Archaeen, sowie deren Habitat, Halotoleranz und Sauerstoffmetabolismus herstellen. Domänentypische Enzyme der Archaeen spielen zumeist eine Rolle beim Aufbau strukturentypischer Phospholipide der Archaeen-gemeinen Zellmembran oder stehen im Kontext mit der Proteinfaltung und Strukturhaltung von DNA und RNA unter den harschen Bedingungen der Umgebungen, in denen sie existieren. Gesondert zu erwähnen ist die Methanogenese der Archaeen, die mit einigen typischen Enzymklassen den Metabolismus dieser Organismen charakterisiert. Ebenfalls sind wiederholt Enzyme mit sehr breiter Substratspezifität typisch für Archaeen annotiert. Die in diesem Kapitel interpretierten Zusammenhänge gilt es nun, mittels weiterführenden Analysen (Kapitel 3.5) zu prüfen, gegebenenfalls zu erweitern und schlussendlich im Gesamtkontext auf Stoffwechselwegebene zu verifizieren (Kapitel 3.5).

3.5 Die Enzym-Matrizen (Metadaten)

Durch das Hinzuziehen weiterer Annotationsquellen, die der EnzymeDetector bietet, können die Zuordnungen von Enzymen mittels einer reinen BLAST-Suche (Kapitel 3.4) in einigen Fällen untermauert oder auch widerlegt werden. Der Vorteil der „optimalen Annotation“ (Kapitel 2.9) kommt zum Tragen und die Interpretationen zu den Archaeen (Kapitel 3.4) können gefestigt werden. Ebenso können Vorhersagen zu enzymatischen Funktionen, die auf Basis der Sequenz-homologie nicht zu erkennen sind, getroffen werden (BrEPS, BRENDA, AMENDA) (Kapitel 1.6).

Archaeentypische Enzyme des Sauerstoffmetabolismus

Die Tabelle 13 listet exemplarisch Enzyme der Archaeen auf, die nach diesem Verfahren ebenfalls, bzw. insbesondere zusätzlich zu Kapitel 3.4 eine Verknüpfung mit den einzelnen Metadaten nachweisen können. Betrachtet man die Gegenüberstellung des Sauerstoffmetabolismus, werden in den Daten insgesamt 77 Enzyme aufgeführt, die mindestens 50% mehr Anteile an der Gesamtgruppe der aeroben Archaeen vorweisen. 28 Enzyme, die in einem zu 50% größeren Anteil der anaeroben Vertreter annotiert sind, können ebenfalls hervorgehoben werden. Neben Enzymen, die zum Beispiel sauerstoffabhängig den transmembranen Protonentransport katalysieren (*ubiquinol oxidase (H⁺-transporting)*, (EC 1.10.3.10) [280]), können auch weitere aerobtypische Funktionen wie die Oxidation spezifischer Fettsäuren (*acyl-CoA oxidase*, (EC 1.3.3.6) [281]) ausgemacht werden. Der besondere Aufbau der Zellmembranen, bei Archaeen vornehmlich durch den Einbau verzweigter Isoprenoideinheiten, (Mevalonat Metabolismus, EC 4.1.1.33 [282]), kann auch hier in Zusammenhang mit den aeroben Vertretern gebracht werden (Kapitel 3.4). Ebenfalls wird die Notwendigkeit des Sauerstoffs durch die typisierte Biosynthese eines Lipoyl Cofaktors deutlich (*lipoyl synthase*, (EC 2.8.1.8) [283]). Mit dessen Hilfe werden bei Archaeen und vor allem in den Mitochondrien der Eukaryoten Liponsäuren synthetisiert. Diese wiederum agieren vornehmlich bei oxidativen Decarboxylierungen zum Transfer von Wasserstoff- und Acylgruppen und stellen ein Bindeglied zwischen der Glycolyse und dem TCA-Zyklus dar. [284] Des Weiteren können Liponsäuren ein biochemisches Redoxsystem darstellen und als Radikalfänger und Antioxidans fungieren. [285] Weiterhin als interessant erweist sich die Carboxymethylenebutenolidase (EC 3.1.1.45). Durch die Hydrolyse von Carboxylestern spielt dieses Enzym eine tragende Rolle beim Abbau oft toxischer Chloroaromaten. [286–288] Einer genauen weiterführenden Untersuchung sollten die Zusammenhänge mit der Bildung verschiedener Antibiotikavorstufen bei den aeroben Archaeen unterzogen werden (EC 3.5.1.11: Penicillin [22], EC 3.5.4.16: preQ₀ [21]).

3 Ergebnisse und Diskussion

Tabelle 13: Metadatentypische Enzyme der Archaeen. Häufig zugeordnete Enzyme werden den Metadaten entsprechend gegenübergestellt. Dementsprechend wird entweder die Differenz der prozentualen Anteile gegenüber allen Archaeen aufgelistet (aerob/anaerob) oder der Prozentbereich der Anteile in Bezug auf die Organismengruppe und den jeweiligen Metadaten dargestellt (pH-Toleranz (acidophil-sauer, alkaliphil-alkalisch), Temperaturoptimum (psychrotolerant, thermophil), Habitat (terrestrisch, heiße Quellen, vulkanisch, Schwefelquellen, Schmutzwasser), Halotoleranz (gering (1-6%NaCl), extrem (>15%NaCl))). Die Funktionen und Namen der Enzyme sind, wenn keine weitere Quelle angegeben ist, aus BRENDA übernommen.

Sauerstoffmetabolismus: aerob (a), anaerob (n)		
EC-Nummer	Differenz %(a) - %(n)	Funktion/Namen [20]
EC 1.3.3.6	77-8=69	Fettsäurenoxidierung [281]
EC 1.3.8.4	84-11=73	<i>L-Leucine Degradation</i>
EC 1.10.3.10	55-1=54	Protonenpumpe bei der Atmung [280]
EC 2.8.1.8	75-5=70	Biosynthese des Lipoyl Cofaktors [283]
EC 3.1.1.45	66-6=60	Abbau von Chloroaromaten [286]
EC 3.4.23.42	57-3=54	Hydrolyse von Peptidbindungen
EC 3.5.1.11	64-5=59	Biosynthese von Antibiotika (Penicillin) [22]
EC 3.5.4.16	70-8=62	preQ ₀ Biosynthese, Antibiotika [21]
EC 4.2.1.68	59-6=53	Fructose und Mannose Metabolismus
EC 4.1.1.33	84-1=83	Mevalonat Metabolismus, Aufbau IPP [282]
EC 4.1.3.1	61-0=61	Glyoxylate Cycle, Bypass TCA Cycle [21]
EC 5.5.1.7	57-5=52	Abbau von 3,5-Dichlorocatechol
EC 1.2.99.5	5-71=-66	<i>cobalt-precorrin-5B (C1)-methyltransferase</i>
EC 1.4.1.21	2-63=-61	NAD Metabolismus
EC 1.5.98.1	2-62=-60	Methanmetabolismus
EC 1.15.1.2	0-68=-68	Superoxid Reduktase
EC 1.17.4.2	0-51=-51	Purinmetabolismus
EC 1.18.6.1	2-56=-54	Stickstoffassimilation
EC 2.3.1.169	2-54=-52	Methanmetabolismus
EC 2.5.1.10	36-90=-54	<i>farnesyl diphosphate synthase</i>
EC 4.1.1.25	20-85=-65	<i>tyrosine decarboxylase</i>
EC 4.1.1.39	9-72=-63	Photosynthese, RuBisCO-like Protein: Methionin Salvage [9]
EC 4.2.1.147	0-61=-61	Abbau/Detoxifizierung von Formaldehyd [289]
pH: sauer (s), alkalisch (al).		
EC-Nummer	%-Bereich	Funktion/Name [20]
EC 2.5.1.16	s, 90-100	<i>spermidine synthase</i>
EC 2.1.1.148	s, 80-89	Pyrimidinmetabolismus
EC 4.1.1.50		<i>adenosylmethionine decarboxylase</i>
EC 4.2.3.4		<i>3-dehydroquinate synthase</i>
EC 2.8.1.6	al, 90-100	<i>biotin synthase</i>
EC 2.3.3.10		<i>hydroxymethylglutaryl-CoA synthase</i>
EC 2.3.1.30	al, 80-89	<i>serine O-acetyltransferase</i>
Temperatur: psychrotolerant (p), thermophil (t).		
EC-Nummer	%-Bereich	Funktion/Name [20]
EC 2.7.7.68	p, 90-100	<i>2-phospho-L-lactate guanylyltransferase</i>
EC 3.5.4.10		<i>IMP cyclohydrolase</i>
EC 3.5.4.27		<i>methenyltetrahydromethanopterin cyclohydrolase</i>
EC 3.6.3.8		<i>Ca²⁺-transporting ATPase</i>
EC 6.3.2.31		<i>coenzyme F420-0:L-glutamate ligase</i>
EC 1.1.1.93	p, 80-89	<i>tartrate dehydrogenase</i>
EC 1.5.1.5		<i>methylenetetrahydrofolate dehydrogenase (NADP+)</i>
EC 4.1.1.73		<i>tartrate decarboxylase</i>
EC 4.2.1.31		<i>maleate hydratase</i>
EC 1.15.1.2	t, 80-89	<i>superoxide reductase</i>

Habitat: terrestrisch (tr), heie Quellen (hq), vulkanisch (v), Schwefelquellen (sq), Schmutzwasser (sw).		
EC-Nummer	%-Bereich	Funktion/Name [20]
EC 3.1.3.3	tr, 80-89	<i>phosphoserine phosphatase</i>
EC 2.5.1.16	hq, 90-100	<i>spermidine synthase</i>
EC 4.1.1.50		<i>adenosylmethionine decarboxylase</i>
EC 2.5.1.16	v, 90-100	<i>spermidine synthase</i>
EC 4.1.1.50		<i>adenosylmethionine decarboxylase</i>
EC 2.5.1.16	sq, 90-100	<i>spermidine synthase</i>
EC 4.1.1.50		<i>adenosylmethionine decarboxylase</i>
EC 4.2.1.6	sq, 80-89	<i>Entner Doudoroff Pathway, Galactosemetabolismus</i>
EC 4.2.3.4		<i>3-dehydroquinate synthase</i>
EC 6.3.4.18		<i>5-(carboxyamino)imidazole ribonucleotide synthase</i>
EC 1.12.98.1	sw, 90-100	<i>coenzyme F420 hydrogenase</i>
EC 1.18.6.1		Stickstoff Assimilation
EC 2.3.1.101		<i>formylmethanofuran-tetrahydromethanopterin N-formyltransferase</i>
EC 2.8.4.1		<i>coenzyme-B sulfoethylthiotransferase</i>
EC 3.1.3.3		<i>phosphoserine phosphatase</i>
EC 4.1.1.79		<i>sulfofpyruvate decarboxylase</i>
EC 4.2.1.147		<i>5,6,7,8-tetrahydromethanopterin hydro-lyase</i>
EC 2.1.1.245	sw, 80-89	<i>5-methyltetrahydrosarcinapterin: corrinoid/iron-sulfur protein Co-methyltransferase</i>
EC 2.3.1.169		Methan Metabolismus
EC 3.6.3.29		molybdate-transporting ATPase
Halotoleranz: gering (g), extrem (e).		
EC-Nummer	%-Bereich	Funktion/Name [20]
EC 1.3.1.14	g, 90-100	<i>dihydroorotate dehydrogenase (NAD+)</i>
EC 2.4.2.57		<i>AMP phosphorylase</i>
EC 4.1.2.43		<i>3-hexulose-6-phosphate synthase</i>
EC 1.2.99.5	g, 80-89	<i>formylmethanofuran dehydrogenase</i>
EC 1.3.98.1		<i>dihydroorotate dehydrogenase (fumarate)</i>
EC 2.1.1.195		<i>cobalt-precorrin-5B (C1)-methyltransferase</i>
EC 1.1.1.146	e, 90-100	<i>11beta-hydroxysteroid dehydrogenase</i>
EC 1.1.1.307		<i>D-xylose reductase</i>
EC 1.3.8.2		<i>4,4'-diapophytoene desaturase (4,4'-diapolycopene-forming)</i>
EC 1.18.1.3		<i>ferredoxin-NAD+ reductase</i>
EC 4.2.1.8	e, 80-89	<i>mannonate dehydratase</i>
EC 6.2.1.4		<i>succinate-CoA ligase (GDP-forming)</i>

Den Archaeen aus dem anaeroben Milieu knnen typische Enzyme im Zusammenhang mit dem Methanmetabolismus (EC 1.5.98.1, EC 2.3.1.169) und der Stickstofffixierung durch sauerstofflabile Enzyme (*nitrogenase*, (EC 1.18.6.1) [20,290]) zugeordnet werden. Weiterhin wird der Abbau verschiedener Intermediate der Methanogenese durch diverse Enzyme der anaeroben Archaeen katalysiert. Dementsprechend ist zum Beispiel das *formaldehyde-activating enzyme* (EC 4.2.1.147) zum Abbau von teils auch toxischen Formaldehyd-Verbindungen vorrangig bei anaeroben Vertretern annotiert.

Archaeen mit extremem pH-Wert- und Temperaturoptimum

Bei den Archaeen aus Habitaten mit extremen pH-Werten und Temperaturen können katalysierte Synthesewege (EC 2.5.1.16) zum Aufbau verzweigtkettiger Polyamine, wie Spermidin, sowohl zum Schutz der DNA und RNA dienen, als auch die Membranstabilität entsprechend unterstützen (Kapitel 3.4). Alkaliphile Vertreter weisen mit der *biotin synthase* (EC 2.8.1.6) eine enzymatische Schlüsselfunktion des letzten Schrittes der Biosynthese von Biotin, einem einerseits wichtigen Cofaktor zur Übertragung von Kohlenstoffgruppen im Kohlenhydratstoffwechsel und andererseits bedeutenden Regulator der Genexpression [291], auf, die bisher eher bei Bakterien und Eukaryoten nachgewiesen ist. [20,292] Psychrophile Archaeen aus sehr kalten Umgebungen exprimieren mit der *tartrate dehydrogenase* (EC 1.1.1.93) [293] und der *tartrate decarboxylase* (EC 4.1.1.73) [294] Enzyme, die die Verwendung von L-Tartrat, einem natürlich vorkommenden Pflanzenstoff, als Kohlenstoffquelle ermöglichen. [295] Dem Methanmetabolismus der Archaeen in Schwefelquellen sind entsprechende Enzyme typischerweise zuzuordnen (*coenzyme F420 hydrogenase* (EC 1.12.98.1), *CO-methylating acetyl-CoA synthase* (EC 2.3.1.169)). Vertreter, die sehr hohe Salzkonzentrationen dulden, können ebenfalls enzymatische Funktionen aufweisen, die für ihre Gruppe typisch sind. Einige davon können alternative Kohlenstoffquellen wie Xylose (EC 1.1.1.307) [296] oder andere oft pflanzliche Zuckerverbindungen (EC 4.2.1.8) [297] verfügbar machen. Andere Auffälligkeiten wie die *4,4'-diapophytoene desaturase* (*4,4'-diapolyycopene-forming*) (EC 1.3.8.2) werden im Zusammenhang mit der Synthese von Carotinoiden genannt [298]. Carotinoide zählen zu den im Laufe der Evolution der Pflanzen entwickelten sekundären Pflanzenmetaboliten zum Schutz vor zu starker Sonneneinstrahlung. Diese Möglichkeit des UV-Lichtschutzes bei halophilen Archaeen ist noch nicht eingehend untersucht. [299,300]

Marine und nicht marine Habitate der Rhodobacteraceaeen

Die Untersuchung einer zweiten Organismengruppe, der Rhodobacteraceen, kann mit diesem Verfahren ebenso typische Enzyme für den marinen, bzw. nicht marinen Lebensraum, sowie weitere Habitate definieren (Tabelle 14).

Dementsprechend können elf Enzyme herausgestellt werden, die anteilig mindestens zu 50% häufiger in marinen Vertretern annotiert sind. Demgegenüber sind 6 enzymatische Funktionen zahlreicher mit mindestens einer Differenz von 50% bei den nicht marinen Rhodobacteraceen annotiert. Die Unterscheidung der Isolationsart der Organismen nach dem freilebenden Wildtyp und den Laborstämmen kann keine entscheidenden Differenzen aufzeigen. Weiterhin können zu den Habitaten, bzw. Lebensstilklassen Biofilm, Frischwasser und Meereis nur sehr wenige Vertreter zugeordnet werden, so dass eine statistische Aussage mit entsprechender quantitativer Grundlage nicht möglich ist.

Die Rolle einiger auffälliger Enzyme in Verbindung mit den entsprechenden Metadaten ist offensichtlich. Typisch für den marinen Lebensraum, kann zum Beispiel die *mercury(II) reductase* (EC 1.16.1.1) herausgestellt werden. Bei hohen Temperaturen und Salzkonzentrationen ist es mit ihrer katalytischen Funktion möglich, toxisches Quecksilber (Hg^{2+}), das bekannterweise oft in marinen Sedimenten zu finden ist, in weniger schädliche und aktive Verbindungen (Hg^0) zu überführen. [126] Dementsprechend wird das selbige Enzym ebenso häufig im Sediment nachgewiesen. Nicht weniger typisch für marine Habitate ist die Notwendigkeit, sowohl zum Umgang mit halogenen Verbindungen, als auch des metabolischen Auf- und Abbaus selbiger. [301] Ein Beispiel hierfür ist die *(S)-2-haloacid dehalogenase* (EC 3.8.1.2), die den Abbau halogenhaltiger Säureverbindungen katalysiert. [302] Die *nitrile hydratase* (EC 4.2.1.84) erlaubt den Umgang mit und derweilen den Abbau von marinen acetonitrilen Abfallstoffen [303] und ist zu über 60% nur in marinen Vertretern annotiert.

3 Ergebnisse und Diskussion

Tabelle 14: Metadatentypische Enzyme der Rhodobacteraceen. Häufig zugeordnete Enzyme werden den Metadaten entsprechend gegenübergestellt. Dementsprechend wird entweder die Differenz der prozentualen Anteile gegenüber allen Rhodobacteraceen aufgelistet (marin/nicht marin lebend) oder der Prozentbereich der Anteile in Bezug auf die Organismengruppe und den Habitaten/Lebensstilen dargestellt (Organismus gebunden, im Sediment, Erdreich, Wasser lebend). Die Funktionen und Namen der Enzyme sind, wenn keine weitere Quelle angegeben ist, aus BRENDA übernommen.

marin(m)/nicht marin lebend(n):		
EC-Nummer	Differenz %(m) - %(n)	Funktion/Name [20]
EC 1.16.1.1	98-16=82	in vivo Detoxifizierung von Quecksilber, bei hohen Temperaturen und Salzkonzentrationen [126]
EC 2.1.1.172	74-16=58	RNA Modifikation [5]
EC 2.1.1.174	65-4=61	rRNA Methylierung, Verbindung der ribosomalen Untereinheiten, trotz hoher Temperaturen und Osmostress [304,305]
EC 2.1.1.250	69-12=57	Methanmetabolismus
EC 2.7.1.20	82-16=76	Purinmetabolismus
EC 3.8.1.2	97-12=85	Kata-/Anabolismus halogenierter Verbindungen [302]
EC 4.1.2.53	81-24=57	Abbau von Hexosen
EC 4.2.1.6	87-28=59	Entner Doudoroff Pathway, Galactose Metabolismus
EC 4.2.1.39	86-28=58	Entner Doudoroff Pathway, Galactose Metabolismus
EC 4.2.1.84	62-0=62	Abbau von stickstoffhaltigen Verbindungen und Aromaten
EC 4.2.1.108	44-0=44	Ectoin-Biosynthese [306]
EC 5.4.99.20	80-4=76	RNA Modifikation
EC 2.7.7.27	34-84=-50	Glycogenmetabolismus
EC 3.5.1.1	34-92=-58	Aspartat- und Asparaginmetabolismus
EC 3.6.3.34	22-80=-58	ABC Transporter
EC 4.1.1.39	14-72=-58	Photosynthese, RuBisCO-like Protein für Methionin Salvage [307]
EC 4.1.1.98	18-80=-62	Decarboxylierung in Coenzyme Q Biosynthese, Elektronentransport bei der Atmung [308]
EC 4.4.1.22	11-52=-41	Formaldehydabbau
EC 5.4.3.2	5-88=-83	<i>L-Lysine fermentation</i> [21]
Habitat: Organismus-gebunden(o), Sediment(s), Erdreich(e), Wasser(w).		
EC-Nummer	%-Bereich	Funktion [20]
EC 1.16.1.1	o, 90-100	in vivo Detoxifizierung von Quecksilber, bei hohen Temperaturen und Salzkonzentrationen [126]
EC 3.8.1.2		Kata-/Anabolismus halogenierter Verbindungen [302]
EC 1.16.1.1	s, 90-100	in vivo Detoxifizierung von Quecksilber, bei hohen Temperaturen und Salzkonzentrationen [126]
EC 3.8.1.2		Kata-/Anabolismus halogenierter Verbindungen [302]
EC 2.1.1.148	e, 90-100	Pyrimidinmetabolismus
EC 4.1.3.30	e, 80-89	propionate fermentation
EC 4.1.1.96		Biofilmbildung [309]
EC 1.10.3.10		Protonenpumpe bei der Atmung
EC 5.4.3.2		<i>L-Lysine fermentation</i> [7]
EC 1.1.2.7		<i>formaldehyde degradation</i>
EC 3.5.1.1		Aspartat- und Asparaginmetabolismus [5]
EC 1.16.1.1	w, 90-100	in vivo Detoxifizierung von Quecksilber, bei hohen Temperaturen und Salzkonzentrationen [126]
EC 3.8.1.2		Kata-/Anabolismus halogenierter Verbindungen [302]

Methanogene Rhodobacteraceen

Einen interessanten Aspekt bietet die enzymatische Funktion der *trimethylamine-corrinoid protein Co-methyltransferase* (EC 2.1.1.250). Sie ist bisher nur bei den Archaeen der *Methanosarcina* beschrieben und katalysiert den Transfer von Methylgruppen in der Methanogenese. [310] Nach derzeitigem Stand der Forschung wird die aerobe Methanogenese nicht direkt mit den Rhodobacteraceen in Verbindung gebracht. Dennoch liegen Nachweise vor [311], dass beim Shift der Algenblüte-Phase der Ozeane zu den nachfolgenden oligotrophen Bedingungen mit nur wenigen Nährstoffen im Sommer der nur wenig zur Verfügung stehende Stickstoff der limitierende Faktor der Biomasseproduktion ist. Diese Verknappung der Stickstoffquelle wird als Auslöser einer messbaren Methanproduktion im sauerstoffreichen Oberflächenwasser angesehen. Dabei tritt die Methanbildung als Nebenprodukt umgestellter und teils noch zu postulierender Stoffwechselwege der Verwertung von Phosphaten und anderen methylierten Verbindungen auf. [311] Zumindest die aerobe und anaerobe Fixierung von Kohlenstoffdioxid als weiteres mögliches Substrat zur Methanogenese kann bei einigen Vertretern nachgewiesen werden (z.B. *Rhodobacter sphaeroides*, [312]).

Die ectoine synthase der marinen Rhodobacteraceen

Ebenso zu diskutieren ist die ausschließliche Annotation der *ectoine synthase* (EC 4.2.1.108) bei marinen Rhodobacteraceen. Mit Hilfe dieser Synthase kann die Kondensationsreaktion eines Diaminobutanoates zum Ectoin katalysiert werden. Ectoin stellt nachweislich sowohl eine Möglichkeit zur Reaktion auf osmotischem Stress als auch einen Schutzfaktor gegen UV-Strahlung bei Mikroorganismen dar. [306,313]

Fermentationsstoffwechsel der Rhodobacteraceen

Für die nicht marinen und im Erdreich lebenden Rhodobacteraceen können Enzyme mit Beteiligung an Fermentationsstoffwechselwegen (EC 5.4.3.2, *L-Lysine fermentation* [21], EC 4.1.3.30, *propionate fermentation* [21]) und dem Abbau von Formaldehyd (EC 4.4.1.22, EC 1.1.2.7 [21]) typisiert werden. Fermentationen an sich sind im sauerstofffreien Metabolismus von Mikroorganismen zur Energiegewinnung evolviert. [314,315] Ebenso sind der Abbau und die Weiterverwertung von Formaldehyd typisch für anaerobe und nicht marine Bakterien. [316,317] Enzyme, die im Zusammenhang mit der Photosynthese stehen (*ribulose-bisphosphate carboxylase*, EC 4.1.1.39), bzw. als Elektronen-Carrier dienen (*4-hydroxy-3-polyprenylbenzoate decarboxylase*, EC 4.1.1.98), werden ebenso vorrangig in nicht marinen Vertretern verzeichnet. Dabei ist aber zum Beispiel im Fall der *ribulose-bisphosphate carboxylase* lediglich eine Homologie zum Schlüsselenzym der Kohlenstoffdioxidfixierung des Calvin Zyklus (RuBisCO) bei Pflanzen nachgewiesen. Diese sogenannten „RuBisCO-like“-Enzyme sind nicht in der Lage, dieselben Carboxylase/Oxygenase-Reaktionen wie für die Photosynthese notwendig, zu katalysieren. Ihnen kann jedoch eine Rolle in der Regenerierung von Methionin über einen Salvage-Stoffwechselweg ebenso nachgewiesen werden, wie der evolutive Zusammenhang als Vorreiter der Entwicklung der photosynthetischen Aktivitäten. [307]

Die Notwendigkeit typischer Enzyme der Rhodobacteraceen

Wie in diesem Kapitel beschrieben, können auch enzymatische Funktionen der Rhodobacteraceen in Zusammenhang mit den entsprechenden Habitaten gebracht werden. Dementsprechend können einige Enzymklassen der marinen Vertreter deutlich von den nicht marinen abgegrenzt werden, während weitere Unterschiede bei im Sediment oder Erdreich lebenden Organismen zu differenzieren sind.

Interessanterweise unterscheiden sich die Überrepräsentationen der marin typisierten Enzyme von jenen nicht mariner Habitate dahingehend, dass diese sehr oft in nahe 90% und mehr ihrer Metadatenklasse annotiert werden können und teils gar nicht in der gegenübergestellten Gruppe vertreten sind. Dementgegen sind die enzymatischen Funktionen der nicht marinen Rhodobacteraceen oft nur in weit geringeren Anteilen von 50-80% ihrer Gruppe zu finden, während jedoch auch teils bis zu 30% der marinen Bakterien über gleiche Funktionen verfügen. Damit scheinen sie offensichtlich nicht essentiell notwendig für diese Lebensweise zu sein. Lediglich spezialisierte Teilgruppen verfügen über entsprechende Funktionalitäten. Dies spiegelt jedoch ebenso die größere Vielfalt der Umweltbedingungen und Habitate nicht mariner Vertreter, die dieser Gruppe in diesem Projekt zugeordnet werden, wieder (Kapitel 2.4).

Mit Hilfe des zweiten Verfahrens können kurzum sowohl die Zusammenhänge archaeentypischer Enzyme aus Kapitel 3.4 aufgegriffen, als auch weitere Aspekte erkannt werden. Die Untersuchung einer zweiten Analysegruppe stellt ebenso die Notwendigkeit typisierter enzymatischer Funktionen des an Habitat und Umweltbedingungen angepassten Metabolismus der Organismen heraus. Diese Interpretationen werden nun im folgenden Kapitel im Kontext vollständiger Stoffwechselwege nachvollzogen und eventuelle weitere Auffälligkeiten dargestellt.

3.6 Die Stoffwechselweg-Matrizen (Metadaten)

Die Zuordnungen von Stoffwechselwegen zu den einzelnen Organismen der untersuchten Projektgruppen ermöglichen eine Auswertung direkt auf Ebene des Metabolismus. Dabei werden nur Pathways zugeordnet, deren enzymatische Voraussetzungen definitiv und komplett durch das Vorhandensein der entsprechenden obligatorischen Schlüsselenzyme oder der Katalyse alternativer Reaktionsschrittfolgen offensichtlich erfüllt sind (Kapitel 3.2).

Die in Kapitel 3.3, bzw. Kapitel 3.5 getroffenen Aussagen bezüglich der Interpretation des verfügbaren Enzympools der Archaeen und Rhodobacteraceen können auch durch die Abbildung auf Stoffwechselwegebene gefestigt werden. Zusätzlich stellen sich noch einige metadatentypische Zusammenhänge heraus, die weitere Aspekte darlegen.

Archaeentypische Stoffwechselwege

Dementsprechend sind in Tabelle 15 zugeordnete Stoffwechselwege der Archaeen herausgegriffen, deren Auftreten im Zusammenhang mit dem Sauerstoffmetabolismus, dem pH-Wert, der Temperatur, der Halotoleranz oder dem Habitat der Vertreter steht. Dabei sind 22 Pathways mit einer Differenz von mehr als fünfzig Prozentanteilen bei aeroben Archaeen annotiert. Demgegenüber stehen 8 typische Biosynthesewege der anaeroben Organismen.

Die β -Oxidation von Fettsäuren (MetaCyc: *fatty acid beta-Oxidation II* (PWY-5136)) ist eigentlich nur bei Pflanzen experimentell nachgewiesen, jedoch codieren knapp 80% der aeroben Archaeen alle benötigten enzymatischen Funktionen auf ihrem Genom. Somit besteht zumindest die funktionelle Grundlage für einen ähnlich angelegten Abbau von Fettsäuren zur Bildung von Acetyl-CoA, welches letzten Endes im TCA-Zyklus weiter zur Energiegewinnung abgebaut wird. [318] Die vollständige Verwertung von Fettsäuren bei den Archaeen ist allerdings noch nicht komplett nachvollzogen und ist Thema einiger Arbeiten. [49,319] Denkbar ist bei den Archaeen sogar eine Umkehr der reversiblen Reaktionen der β -Oxidation unter Verwendung der hier gefundenen Enzyme. Dann können beispielsweise mit demselben Produktionsmechanismus von Acetyl-CoA/Acetoacetyl-CoA aus dem Mevalonatweg (Kapitel 3.4) Fettsäuren als Energiespeicher synthetisiert werden. [320] Ein anderer Stoffwechselweg ermöglicht den aeroben und extrem halotoleranten Archaeen die Bioverfügbarkeit zum Beispiel von Schwefel zum Einbau in Aminosäuren und Coenzyme (MetaCyc: *thiosulfate disproportionation III (rhodanese)* (PWY-5350)). [321]

Der Abbau toxischer Chloroaromaten kann ebenfalls auf Stoffwechselwegebene durch diverse Pathways gegenüber den aeroben Organismen dieser Projektgruppe typisiert werden (Kapitel 3.4) (MetaCyc: *5-chloro-3-Methyl-Catechol degradation* (PWY-6102), *3,4-dichlorotoluene degradation* (PWY-6192)). Die Phosphorylierung von zum Beispiel Threonin und Serin spielt bei den Archaeen eine große Rolle im Zusammenhang mit der Signaltransduktion zellulärer Prozesse. Dies kann vor allem bisher bei aeroben Archaeen gezeigt werden [322] und bietet eine Erklärung von verstärkt auftretenden Synthese- und Abbauwegen von Threonin.

3 Ergebnisse und Diskussion

Tabelle 15: Metadaten typische Stoffwechselwege der Archaeen. Häufig zugeordnete Stoffwechselwege werden den Metadaten entsprechend gegenübergestellt. Dementsprechend wird entweder die Differenz der prozentualen Anteile gegenüber allen Archaeen aufgelistet (aerob/anaerob) oder der Prozentbereich der Anteile in Bezug auf die Organismengruppe und den jeweiligen Metadaten dargestellt (pH-Toleranz (acidophil-sauer, alkaliphil-alkalisch), Temperaturoptimum (psychrotolerant, thermophil), Habitat (terrestrisch, aquatisch, heiße Quellen, vulkanisch, Schwefelquellen, Schmutzwasser), Halotoleranz (gering (1-6%NaCl), extrem (>15%NaCl))). Ist ein Stoffwechselweg laut der „taxonomic range“ von MetaCyc (Kapitel 3.7) nicht in der Projektgruppe der Archaeen zu erwarten, ist er mit (*) markiert. Alle beschriebenen Namen, bzw. Funktionen der genannten Stoffwechselwege stammen aus der MetaCyc-Datenbank.

Sauerstoffmetabolismus: aerob (a), anaerob (n)	
Pathway [21]	Differenz %(a) - %(n)
fatty acid β -Oxidation II (*)	77-7=70
thiosulfate disproportionation III (rhodanese) (*)	89-18=71
glycine cleavage	82-14=68
glycine biosynthesis II (*)	82-14=68
L-threonine degradation V (*)	89-22=67
L-threonine biosynthesis	98-67=61
superpathway of L-threonine biosynthesis	91-31=60
5-chloro-3-methyl-catechol degradation (*)	57-5=52
3,4-dichlorotoluene degradation (*)	57-5=52
L-proline degradation	61-10=51
L-arginine biosynthesis II	2-62=-60
faktor 420 biosynthesis	2-61=-59
beta-alanine biosynthesis III	34-92=-58
L-cysteine biosynthesis II	2-59=-57
methyl-coenzyme M reduction to methan	2-56=-54
nitrogen fixation I (ferredoxin)	2-56=-54
farnesyl diphosphate biosynthesis	36-89=-53
geranyl diphosphate biosynthesis	39-89=-50
L-asparagine biosynthesis I (*)	30-77=-47
pH: sauer (s), alkalisch (al).	
Pathway	%-Bereich
spermidine biosynthesis I	s, 80-89
parathion; paraoxon degradation (*)	s, 41, aber ausschließlich
L-asparagine biosynthesis I (*)	al, 90-100
L-cysteine biosynthesis I	al, 80-89
Temperatur: psychrotolerant (p), thermophil (h).	
Pathway	%-Bereich
faktor 420 polyglutamylation	p, 90-100
L-leucine biosynthesis	
L-valine biosynthesis	
spermidine biosynthesis I	t, 90-100
Habitat: heiße Quellen (hq), vulkanisch (v), Schwefelquellen (sq), Schmutzwasser (sw).	
Pathway	%-Bereich
spermidine biosynthesis I	hq, 90-100
spermidine biosynthesis I	v, 90-100
spermidine biosynthesis I	sq, 90-100
taxiphyllin bioactivation (*)	sq, 70-89
faktor 420 polyglutamylation	sw, 90-100
faktor 420 biosynthesis	
L-leucine biosynthesis	
L-valine biosynthesis	
L-arginine biosynthesis II (acetyl cycle)	

nitrogen fixation I (ferredoxin)
tetrapyrrole biosynthesis I (from glutamate)
thiamine salvage II
methanogenesis

sw, 80-89
 sw, 50-79,
 aber ausschließlich

Halotoleranz: gering (g), extrem (e).

Pathway	%-Bereich
<i>D-mannose degradation</i>	g, 90-100
<i>GDP-mannose biosynthesis</i>	g, 80-89
<i>FeMo cofactor biosynthesis</i>	
<i>thiosulfate disproportionation III (Rhodanese)</i>	e, 90-100
<i>cyanide degradation (*)</i>	
<i>glycine biosynthesis II</i>	e, 80-89
<i>glycine cleavage</i>	

Bei den anaeroben Archaeen und in Schmutz- und Abwasser lebenden Vertretern zeigen sich dem annotierten Enzympool zufolge entsprechende Stoffwechselwege des Methanmetabolismus vorrangig bis ausschließlich in diesen Organismengruppen (*factor 420* Synthese/Umwandlungen und Methanogenese). Die Biosynthese von Farnesyl- und Geranyl-Diphosphat sind als Intermediate des Mevalonat-Pathways im Zusammenhang mit dem Aufbau von Isoprenoideinheiten für die archaeentypische Zellmembran zu erwähnen (Kapitel 3.4, Kapitel 3.9). Deren Synthese scheint funktionell den Daten zufolge vorrangig bei anaeroben Vertretern annotiert. Allerdings, so deuten es die Daten ebenfalls mit knapp 40% der Gruppenanteile an, können entsprechende experimentelle Beobachtungen fortwährend auch bei aeroben Archaeen gemacht werden. [261,323] Ebenfalls im Hinblick auf die spezifische Membranstruktur kann die *spermidine biosynthesis I* (MetaCyc: BSUBPOLYAMSYN-PWY) typisch für acidophile und thermophile Archaeen aus den temperaturextremen Habitaten (vulkanisch, Schwefelquellen, heiße Quellen) genannt werden (Kapitel 3.4, Kapitel 3.11).

Ausschließlich einige acidophile Vertreter sind rein enzymatisch betrachtet dazu in der Lage, die außerordentlich toxischen Substanzen Parathion und ein oxidiertes Derivat Paraoxon abzubauen. Diese wurden und werden verbreitet als Insektizide eingesetzt, da sie ihre toxische Wirkung nur gegen Insekten und Säugetiere entfalten. [324] Es kann allerdings bei einigen Bakterien und Archaeen gezeigt werden, dass eben diese Stoffe im Labor als einzige Kohlenstoffquelle zur Energiegewinnung und zum Überleben ausreichen. [325,326] Der evolutionäre Hintergrund ist allerdings derweilen noch Gegenstand publizierter Interpretationen. Es wird diesbezüglich auf Grund funktioneller und struktureller Verknüpfungen eine parallele Evolution der Paraoxinasen mit Enzymen der Klasse der Laktonasen zur Hydrolyse von Lactonen (Kapitel 3.4: Entner-Doudoroff-Weg bei acidophilen Archaeen) postuliert. [326,327]

Vorrangig bei Archaeen aus Schmutz- und Abwasser dient ein Salvage-Pathway (MetaCyc: thiamine salvage II (PWY-6897)) der Wiederverwendung des Thiamin-Diphosphat, besser bekannt als Vitamin B1. [21] Dies ist ein essentieller Cofaktor vieler Enzyme des Energiestoffwechsels, wie zum Beispiel unter anderem der Pyruvat Dehydrogenasen und Decarboxylasen. [328,329] Die komplexe *de novo*

Biosynthese bei Archaeen des Thiamin-Diphosphat selbst ist immer noch kaum verstanden und obliegt weiterer Forschung. [330]

Die Bioaktivierung des Taxiphyllin (MetaCyc: PWY-7089) wird mit Hilfe der β -glucosidase (EC 3.2.1.21) katalysiert und ist nur bei Vertretern aus dem Reich der Pflanzen zur Produktion von Cyanwasserstoff (Blausäure) bekannt. Dort dient es vornehmlich dem Fraßschutz der Pflanzensamen und Blätter. [21,331] Die am häufigsten bei Archaeen aus Schwefelquellen annotierte *beta-glucosidase* weist eine sehr breite Substratspezifität auf und kann einige Zuckerderivate hydrolysieren (β -D-galactosides, α -L-arabinosides, β -D-xylosides, β -D-fucosides). [20] Damit ist diese Funktionalität bei diesen Archaeen-Vertretern höchstwahrscheinlich vielmehr am Abbau von Cellulose unter hohen Temperaturen zusammen mit einem weiteren Enzym (*cellulase*, EC 3.2.1.4) zu β -D-Glucose beteiligt. [332]

Mit steigender Halotoleranz der Archaeen bis hin zu den extrem halophilen wird der Abbau von Cyaniden mittels der *L-3-cyanoalanine synthase* (EC 4.4.1.9) [21] häufiger annotiert. Dieses Enzym ist nur aus Eukaryoten und Bakterien bekannt [20] und kann in bifunktionalen Zusammenhang mit einer *cysteine synthase* (EC 2.5.1.47) gebracht werden. [333] Der Aufbau von Cystein bei Archaeen [334,335] wird verstärkt zur Bildung von Disulfidbindungen zur Stabilisierung von Proteinen unter Extrembedingungen beobachtet. [218,219]

Rhodobacteraceentypische Stoffwechselwege

Die Vielfaltigkeit der Archaeen ist bekannt und bedingt durch die Anpassung an verschiedenste extreme Habitate ist ein breit aufgestellter Metabolismus der Domäne, wie gezeigt, zu erwarten (Kapitel 1.7). Eine annähernd große metabolische Diversität kann bei den Rhodobacteraceen, wohlgermerkt einer Organismengruppe weit verbreitet in allen Weltmeeren und nicht marinen Habitaten (Kapitel 1.8), vorausgesetzt werden. [61,63,77]

Den Vertretern der Rhodobacteraceen können insgesamt 563 eindeutige Stoffwechselwege zugeordnet werden, wobei davon 206 in über 50% der Gruppe funktionell möglich sind. Dabei lassen sich auch hier Verknüpfungen zu den Metadaten erkennen. Die getroffenen Aussagen und Interpretationen auf Grundlage des Enzymools (Kapitel 3.5: Rhodobacteraceen) können auch auf Stoffwechselwegebene bestätigt werden. Zusätzlich fallen noch Häufigkeiten einiger Beispiele, vor allem typisch für marine Vertreter den nicht marinen Lebensräumen gegenübergestellt, auf. Diesbezüglich sind der Tabelle 16 drei Stoffwechselwege, die zu 50% häufiger mit marinen Organismen im Vergleich zu den nicht marinen Rhodobacteraceen verknüpft sind, zu entnehmen. Vier können auf diese Weise entsprechend den nicht marinen Vertretern zugeordnet werden. Diese werden im Folgenden diskutiert.

Tabelle 16: Metadaten-typische Stoffwechselwege der Rhodobacteraceen. Häufig zugeordnete Stoffwechselwege werden den Metadaten entsprechend gegenübergestellt. Dementsprechend wird entweder die Differenz der prozentualen Anteile gegenüber allen Rhodobacteraceen aufgelistet (marin/nicht marin lebend) oder der Prozentbereich der Anteile in Bezug auf die Organismengruppe und den Habitaten/Lebensstilen dargestellt. (Organismus gebunden, im Sediment, Erdreich, Wasser lebend) Ist ein Stoffwechselweg laut der „taxonomic range“ von MetaCyc (Kapitel 3.7) nicht in der Projektgruppe der Rhodobacteraceen zu erwarten, ist er mit (*) markiert. Alle beschriebenen Namen, bzw. Funktionen der genannten Stoffwechselwege stammen aus der MetaCyc-Datenbank.

marin (m)/nicht marin lebend (n):	
Pathway	Differenz %(m) - %(n)
<i>adenine/adenosine salvage VI</i>	82-16=66
<i>indole-3-acetate biosynthesis IV</i>	57-0=57
<i>acrylonitrile degradation I</i>	57-0=57
<i>ectoine biosynthesis</i>	37-0=37
<i>rubisco shunt (*)</i>	11-72=-61
<i>NAD phosphorylation and dephosphorylation</i>	20-80=-60
<i>L-asparagine degradation I</i>	34-92=-58
<i>glycogen biosynthesis I (ADP-D-glucose)</i>	34-84=-50
Habitat: Organismus-gebunden(o), Sediment(s), Erdreich(e), Wasser(w).	
Pathway	%-Bereich
<i>S-methyl-5'-thioadenosine degradation II</i>	o, 90-100
<i>S-methyl-5'-thioadenosine degradation II</i>	s, 90-100
<i>pyrimidine deoxyribonucleotides de novo biosynthesis III</i>	e, 80-89
<i>L-asparagine degradation I</i>	
<i>L-selenocysteine biosynthesis I</i>	e, 60-79
<i>adenine/adenosine salvage VI</i>	w, 90-100
<i>S-methyl-5'-thioadenosine degradation II</i>	w, 80-89

Die vollständige Ectoin-Biosynthese ist auch auf Stoffwechselwegebene ausschließlich den marinen Bakterien zuzuordnen (Kapitel 3.5: Die *ectoine synthase* der marinen Rhodobacteraceen). Dieser Osmoprotektant und Schutzfaktor gegen UV-Licht vornehmlich halophiler Bakterien [306] ermöglicht folglich die direkte Unterscheidung der Metadatenkategorien bezüglich des marinen Lebensraumes. Ebenso haben sich ein Recyclingstoffwechselweg zur Regenerierung von Adenosin-Nukleotiden (MetaCyc: *adenine/adenosine salvage* (PWY-6619)) [21] und ein Weg, um den Abbau mariner, acrylonitriler/ acetonitriler Abfallstoffe sowohl als Kohlenstoff-, als auch als Stickstoffquelle zu nutzen (MetaCyc: *acrylonitrile degradation I* (PWY-7308)) [21,336], bei den marinen Bakterien entwickelt. Die Herstellung des Phytohormons Indol-3-acetat (MetaCyc: *indole-3-acetate biosynthesis IV* (PWY-5025)) dient den mit Pflanzen assoziierten Bakterien dazu, die Stoffproduktion der Wirtspflanzen ihrem Bedarf entsprechend zu induzieren. [337,338] Dieser funktionelle Faktor kann folglich den marinen Bakterien der Rhodobacteraceen im Zusammenleben mit Phytoplankton (Kapitel 1.8) ermöglichen, benötigte Stoffe als Kohlenstoff-, Stickstoff- und letztlich Energiequelle vom symbiontischen Partner einzufordern. [62,339]

Die NAD Phosphorylierung, bzw. Dephosphorylierung (MetaCyc: *NAD phosphorylation and dephosphorylation* (NADPHOS-DEPHOS-PWY)) wird zu großen Teilen nur den nicht marinen Organismen zugeordnet. Der Grund hierfür ist in der bidirektionalen Ausrichtung dieses Stoffwechselweges zu finden.

3 Ergebnisse und Diskussion

Während die NAD-Kinase (EC 2.7.1.23) in allen Vertretern, ob marin oder nicht, annotiert ist und die Übertragung eines Phosphatrestes katalysiert, mangelt es bei den marinen Rhodobacteraceen an einer Annotation für eine entsprechende Phosphatase in entgegengesetzter Richtung (EC 3.1.3.1, 21% marine Vertreter, 68% nicht marine Vertreter, EC 3.1.3.2, 1% marine Vertreter, 16% nicht marine Vertreter). Die Untersuchung und Charakterisierung von NADP⁺ Phosphatasen ist allerdings trotz ihrer entscheidenden Rolle im Zusammenhang mit diesem Cofaktor weiterhin verbreiteter Gegenstand einiger Arbeiten [340–342] und sollte auch in den Analysen der Rhodobacteraceen entsprechend berücksichtigt werden.

Einige spezifische Aminosäureabbau und Synthesewege können den einzelnen Habitaten zugeordnet werden. Allerdings ist hier ein eindeutiger Zusammenhang dieser Verknüpfungen nicht zu erkennen und bedarf weiterführender Untersuchungen.

Zusammenfassung

Alles in allem ist es von hoher Bedeutung, die entsprechenden enzymatischen Funktionen auf Ebene der Stoffwechselwege zu aggregieren und in Verbindung mit den Umweltbedingungen und Metadaten der Organismen zu analysieren. Das entwickelte Programm (Kapitel 3.2) schafft hier die essentielle und verlässliche Grundlage dieser Zuordnungen. Dementsprechend können die Interpretationen auf Enzymebene bestätigt werden. Ergänzend können aber auch weitere interessante Aspekte im Sinne der Anpassungen der extremophilen Archaeen aufgezeigt werden und beispielsweise ausschließlich den marinen Rhodobacteraceen vollständige Stoffwechselwege, wie die Ectoin-Biosynthese, zugeordnet werden. Dieses Verfahren ist ohne Zweifel allerdings direkt von der Quantität und der Qualität der zuzuordnenden Stoffwechselwege (MetaCyc) abhängig. Eine strikte Differenzierung von katabolischen und anabolischen Reaktionsketten ist eine Voraussetzung, die es einzuhalten gilt. Ein entsprechendes negativ Beispiel ist die Glycolyse zum Abbau von Glucose-6-Phosphat zum Endprodukt Pyruvat (MetaCyc: GLYCOLYSIS). Die Datenbank MetaCyc bezieht hier allerdings obligatorisch das Enzym *phosphoenol pyruvate synthetase* (EC 2.7.9.2) für die Rückreaktion des Pyruvats zum Phosphoenolpyruvate mit ein. Dieses steht dann zum Beispiel Fermentationsstoffwechselwegen anaerober Organismen zur Verfügung. [21] Der Vorgang entspricht allerdings nicht der Glycolyse und ist eher der Gluconeogenese zuzuordnen. [13,20] Für die Produktion des Endproduktes Pyruvat der Glycolyse darf die enzymatische Funktion der EC 2.7.9.2 folglich nicht als obligatorisch angesehen werden. Diese Problematik wurde in diesem expliziten Fall in Anbetracht der Essenz der Glycolyse im Metabolismus bei der Erstellung des Programmes zur Erkennung von Schlüsselenzymen der Stoffwechselwege (Kapitel 3.2) berücksichtigt. In weiterführenden Analysen muss dem allerdings datenbankweit Rechnung getragen werden. Dementsprechend sollten Möglichkeiten gefunden werden, um beispielsweise

die Gruppierungen von Reaktionen der MetaCyc-Datenbank kontextbasiert weiter aufzugliedern, so dass beispielsweise Auf- und Abbau von Metaboliten ausschließlich getrennt betrachten werden und keine Rolle bei der Differenzierung von alternativen Funktionen spielen. Des Weiteren erfüllen diese Bedingungen aktuell aufgebaute Datenquellen wie die BRENDA-Stoffwechselwegkarten (*pathway maps*) [20] grundsätzlich durch die manuelle Überprüfung und Zusammenstellung der Reaktionsfolgen. Der entscheidende Faktor ist hier wiederum die Quantität an Stoffwechselwegen verschiedener Organismen und Domänen, die allerdings mit Sicherheit über die Zeit wachsen wird. Die vorherige Analyse auf Basis enzymatischer Funktionen ist allerdings dessen ungeachtet von ebenso großer Bedeutung, da nicht zwangsweise alle Enzymklassen Reaktionen des Stoffmetabolismus im Sinne der Produktion von Energie und letztlich Biomasse katalysieren. Regulatorische Funktionen und beispielsweise Schutz- und Strukturmechanismen von Proteinen und vor allem DNA und RNA, spielen bei der Verknüpfung zu den vorherrschenden Bedingungen der Organismenhabitate ebenfalls, wie in Kapitel 3.4 und Kapitel 3.5 erwähnt, eine tragende Rolle.

3.7 Die gruppentypischen Lücken in Stoffwechselwegen

Die Betrachtung wiederkehrender Enzymlücken der Analysegruppen soll Hinweise auf eventuelle gruppeneigene Funktionen geben, insofern die verbleibenden enzymatischen Anforderungen des Stoffwechselweges abgedeckt werden. Viele Pathways sind allerdings für individuelle taxonomische Gruppen schon allein durch die bisher bekannte und damit zu erwartende sogenannte „taxonomic range“ auszuschließen. Diese gibt taxonomische Gruppen im Sinne von Domänen, Klassen und Familien an, in denen der betrachtete Stoffwechselweg auf Grund der experimentellen Nachweise in einzelnen Spezies zu vermuten ist. [21] Jedoch sind auch Wege aus den Domänen der Bakterien und Archaeen in den Projektgruppen gehäuft durch die fehlende Annotation einzelner Enzymfunktionen ausgeschlossen. Diese können entweder in den Rhodobacteraceen, bzw. Archaeen tatsächlich nicht vorhanden sein oder die Evolution entwickelte hier bisher nicht klassifizierte Enzyme innerhalb der Projektgruppe. Diese müssen sich auf der Ebene der Sequenzhomologie stark differenzieren.

Sind einzelne Enzymklassen mit einer obligatorischen Schlüsselfunktion in der Reaktionskette in den analysierten Organismen nicht annotiert, wird der entsprechende Stoffwechselweg infolgedessen nicht zugeordnet, gleichwohl ob weitere enzymatische Funktionen abgedeckt sind oder nicht. Dieses Verfahren auf eine ganze Domäne (Archaea) angewendet, hat zur Folge, dass häufig vorkommende Enzyme, die wie beschrieben einen Pathway ausschließen, zumeist auch aus anderen Domänen stammen. Daraus resultiert, dass oft nicht nur einzelne Reaktionsschritte fehlen, sondern komplette Teilläste der Pathways. Demzufolge gestaltet sich die Einordnung in gruppentypische Lücken in Stoffwechselwegen, die die Domäne dennoch zu schließen vermag, als sehr komplex.

Häufige Stoffwechselweglücken der Rhodobacteraceen

Die Tabelle 17 führt diverse entsprechende Enzyme auf. Einige Vertreter der Rhodobacteraceen haben eine Resistenz gegenüber Polypeptid-Antibiotika wie Polymyxin nachgewiesen (zum Beispiel *Wenxinia marina* [343]). Polymyxin besitzt ganz ähnlich zu Zellmembranen einen polaren Aufbau aus entsprechenden Aminosäuren und endständigen hydrophoben Fettsäuren. Aufgrund dessen werden sie in die Zellmembran eingelagert und manipulieren deren Permeabilität im negativen Sinne. [344] Tatsächlich können Annotationen für die meisten enzymatischen Funktionen des bekannten Stoffwechselweges zur Resistenz gegenüber Polymyxin (MetaCyc: *polymyxin resistance* (PWY0-1338) [21]) in fast allen Vertretern ausfindig gemacht werden (EC 1.1.1.305, EC 2.6.1.87, EC 2.1.2.13). Jedoch fehlen Transferasefunktionen für zwei finale Schritte ausgehend von *UDP-4-deoxy-4-formamido-β-L-arabinopyranose* (EC 2.4.2.43, EC 2.4.2.53). Da bei keinem Vertreter dieser Weg bisher experimentell nachvollzogen ist, können nur zwei Vermutungen aufgestellt werden. Entweder wird dieses Substrat durch entsprechende Enzyme auf einem anderen Weg weiter verstoffwechselt oder es sind weitere entsprechende katalytische

Transferasefunktionen evolviert. Da für diese Substrate keine entsprechend geeigneten Enzyme bekannt sind [20] und Laborversuche die Notwendigkeit zumindest der Formyltransferase (EC 2.4.2.53) [345] festigen, ist von Letzterem auszugehen. Dieser Zusammenhang muss allerdings weiter und vor allem experimentell verifiziert werden.

Tabelle 17: Gruppentypische Pathway-Lücken. Es werden Enzyme und ihre Namen aufgelistet, die gehäuft keine Annotationen in den Projektgruppen Rhodobacteraceen und Archaeen nachweisen. Ferner werden Stoffwechselwege zugeordnet, in denen diese Enzyme eine Schlüsselfunktion für einen obligatorischen Reaktionsschritt katalysieren. Enzyme, die hingegen in der Projektgruppe annotiert sind, werden in Klammern stehend nach dem Pathway aufgeführt.

Rhodobacteraceae:		
Enzym	Name	Pathway (vorhandene Enzyme) [21]
EC 2.4.2.43/53	4-amino-4-deoxy-L-arabinsyltransferase, 4-deoxy-4-formamido-L-arabinoasetransferase	Polymyxin Resistenz (EC 1.1.1.305, EC 2.6.1.87, EC 2.1.2.13)
EC 4.1.2.50	6-carboxytetrahydropterin synthase	preQ0 Biosynthese, Antibiotika Biosynthese (EC 3.5.4.16, EC 4.3.99.3, EC 6.3.4.20)
EC 4.2.1.79/117	2-methylcitrate dehydratase	2-methylcitrate Zyklus I und II: Propanoat/ Propanyl-CoA zu Pyruvat/ Succinat (EC 6.2.1.17, EC 2.3.3.5, EC 4.2.1.99, EC 4.1.3.30)
EC 5.4.99.16	trehalose synthase	Trehalose Biosynthese IV. (Projektgruppe verfügt über Trehalose Biosynthese I/V)
EC 6.3.1.1, EC 6.3.5.4, EC 6.3.5.6	asparagin synthase, asparaginyI-tRNA synthase	L-Asparagin Biosynthese I, II, III von L-Aspartat (EC 3.1.1.29)
Archaea:		
Enzym	Name	Pathway (vorhandene Enzyme) [21]
EC 3.5.3.8	formimidoylglutamase	Histidin Degradation I (selten EC 4.3.1.3, EC 4.2.1.49, EC3.5.2.7 annotiert)
EC 4.1.3.38	aminodeoxychorismate lyase	4-aminobenzoate Biosynthese (EC2.6.1.85) Tetrahydrofolat Metabolismus [20]-

Die Trehalose-Biosynthese bei Rhodobacteraceen

Die Trehalose-Biosynthese verweist auf ein weiteres Beispiel für eine Lücke in den Stoffwechselwegen der Rhodobacteraceen. Trehalose ist ein Disaccharid, welches bei vielen Organismen als osmotisch aktive Substanz Osmostress entgegenwirkt [346], als Thermoprotektant dient [347] oder einfach als Kohlenstoffquelle für Metabolite [348] zum Beispiel der Zellwand verwendet wird [349]. Bei der Synthese nach dem Stoffwechselweg der *trehalose biosynthesis IV* (MetaCyc: PWY-2622) wird Trehalose ausgehend von Maltose mit der maltose alpha-D-glucosyltransferase (EC 5.4.99.16) isomerisiert. Dieses Enzym lässt sich allerdings nur bei unter 20% der untersuchten Rhodobacteraceen annotieren. Für diese phylogenetische Gruppe können durch die vorliegenden Daten jedoch zwei alternative Stoffwechselwege ausgehend von anderen Substraten, deren Enzyme auch über die Gensequenzhomologie nachgewiesen werden können, ausfindig gemacht werden (MetaCyc: *trehalose biosynthesis I* (TRESYN-PWY) [81], *trehalose biosynthesis V* (PWY-2661)).

Häufige Stoffwechselweglücken der Archaeen (Tabelle 17)

Die *formimidoylglutamase* (EC 3.5.3.8) führt durch die mangelnde Annotation unter den Archaeen sehr oft zum Ausschluss einer Variante des Histidin-Abbauweges (MetaCyc: *histidine degradation I* (HISDEG-PWY)), während alle weiteren enzymatischen Funktionen (EC 4.3.1.3, EC 4.2.1.49, EC 3.5.2.7) zumindest bei einigen wenigen Vertretern nachgewiesen werden können. Dieser Abbau von Histidin ist bei Bakterien weit verbreitet und dient als Kohlenstoff- (Format/Formamid), Stickstoff- (Ammoniak) und Energiequelle (Glutamat). [350] Auch hier hat sich vermeintlich ein alternativer Weg bei den Archaeen und den Eukaryoten im Laufe der Evolution entwickelt (MetaCyc: *histidine degradation III* (PWY-5030)). Ausgehend vom Substrat des Reaktionsschrittes formimino-L-Glutamat werden mit Hilfe der *glutamate formimidoyltransferase* (EC 2.1.2.5) und Tetrahydrofolat 5-Formiminotetrahydrofolat und ebenfalls L-Glutamat gebildet. [21,351]. Dieses Folat kann dann über weitere Transformationen (EC 4.3.1.4, EC 3.5.4.9) auch als Cofaktor anderen Reaktionen zur Verfügung gestellt werden. Der Stoffwechselweg ist bisher nur für Eukaryoten nachgewiesen. Die Enzyme EC 2.1.2.4, EC 4.3.1.4 und EC 3.5.4.9 können allerdings den Daten zufolge auch bei wenigen Archaeen (*Haloarcula*, *Acidolipifundum*) annotiert werden. Ähnlich der Biosynthese von Histidin scheint die Evolution des Abbaus folglich bei den Archaeen ebenso zumindest den Eukaryoten funktionell analoge Enzyme hervorgebracht zu haben. [352] Da sequenzhomologe Nachweise allerdings nur bei sehr wenigen Vertretern gefunden werden können, bedarf dies weiterer Untersuchungen.

Weiterhin in diesem Zusammenhang ist die *aminodeoxy-chorismate lyase* (EC 4.1.3.38) zu erwähnen. Nur bei sehr wenigen Vertretern (unter 10%) kann sie vorhergesagt werden, wohingegen die Bildung des Substrates katalysiert durch die *aminodeoxychorismate synthase* (EC 2.6.1.85) aus Chorismat bei über 50% der Archaeen annotiert ist. Auch hier kann eine archaeeentypische enzymatische Funktion vermutet werden, was allerdings durch weiterführenden Analysen zu bestätigen ist.

Die Einsichtnahme von auffälligen Lücken in Stoffwechselwegen der Analysegruppen kann also zusammengefasst Hinweise zur weiteren Analyse verschiedener enzymatischer Funktionen geben. Einige Alternativen in Form von einzelnen Enzymen oder vollständigen Stoffwechselwegen können allerdings durch die Recherche in weiterführenden Datenbanken, wie in diesem Kapitel gezeigt, schon vermutet werden.

3.8 Neue gruppentypische Stoffwechselwegansätze

Einige Enzyme kommen in den Analysegruppen gehäuft vor. Insofern kann jeder Organismus der untersuchten Gruppe eine entsprechende enzymatische Funktion im Genom codiert haben. Jedoch werden diese Enzyme teilweise keinem Stoffwechselweg in der MetaCyc-Datenbank zugeordnet. Das folgende Kapitel diskutiert, ob diese Enzyme einem Kontext zusammenhängender Reaktionsfolgen im Sinne der Anpassung der Organismen an ihre Habitate zu unterstellen sind.

Bei den Rhodobacteraceen sind insgesamt 753 Enzyme ohne zugeordneten Pathway zu verzeichnen. 34 davon sind bei 100% der Projektgruppe annotiert. 160 verschiedene Funktionen sind bei 90% und 200 bei 80% der Projektgruppe ohne Zuordnung eines Stoffwechselweges. Der Enzympool der Archaeen hat insgesamt 840 Enzyme ohne Pathway. Fünfzehn sind bei allen untersuchten Vertretern, 30 bei 90% und 55 bei 80% der Projektgruppe annotiert. Die folgende Betrachtung geht der Frage auf den Grund, warum diese Enzyme ohne Zuordnung sind und ob eventuell ein Ansatz eines Stoffwechselweges in diesem Kontext und den entsprechenden Enzymsubstraten postuliert werden kann.

Enzyme von Signal- und Transportwegen

Vorrangig fallen Enzyme mit Funktionen in Signal- und Transport-Pathways auf. Weiterhin werden DNA und RNA modifizierende Enzyme ignoriert, da sie für die Stoffumwandlungen der Organismengruppen hier keine entscheidende Rolle spielen. Diese Enzyme werden in der verwendeten Datenbank keinen metabolischen Stoffwechselwegen zugeordnet. Sie sind jedoch für die Organismen unverzichtbar (Kapitel 3.4: Domänentypische Enzyme). Andere Datenbanken für Stoffwechselwege wie BRENDA und KEGG führen Enzyme dieser Art und Bedeutung mittlerweile zum Beispiel als „nicht Pathway bezogene“ Funktionen gesondert auf, sodass zumindest die Funktion erkenntlich ist.

3 Ergebnisse und Diskussion

Die Tabelle 18 hebt einige Beispiele von Enzymen in den Projektgruppen, die nach diesem Verfahren ausfindig gemacht werden können, hervor.

Tabelle 18: Gruppentypische Enzyme ohne Pathwayzuordnung. Die Tabelle zeigt Enzyme, die in der untersuchten Organismengruppe der Rhodobacteraceen, bzw. Archaeen gehäuft auftreten, in der MetaCyc-Datenbank jedoch keinem Stoffwechselweg zugeordnet werden. Mögliche Funktionen und Ansätze für das Zusammenstellen eines Stoffwechselweges werden entsprechend der enzymatischen Funktion zugeordnet.

Rhodobacteraceae:		
Enzym	Name	Pathway-Ansatz
EC 1.1.1.47	<i>glucose 1-dehydrogenase [NAD(P)+]</i>	Entner Doudoroff Pathway: Alternative für 1. Reaktionsschritt von Glucose zu <i>D-glucono-1,5-lactone</i> [20]
EC 1.8.4.11	<i>peptide-methionine (S)-S-oxide reductase</i>	Vermeidung von Schäden durch oxidativen Stress [353,354]
EC 1.11.1.15	<i>peroxiredoxin</i>	Reaktion auf oxidativen Stress [355]
EC 2.7.7.59	<i>[protein-P1I] uridylyltransferase</i>	Reaktion auf Stickstofflimitierung [356]
EC 2.7.13.3	<i>histidine kinase</i>	Säure- und Osmosensor [357,358]
EC 3.5.1.3	<i>omega-amidase</i>	Aspartat- und Asparaginmetabolismus. Shunt: Asparagin über EC 2.6.1.14 und EC 3.5.1.3 zu Oxalacetat [20]
EC 3.5.1.11	<i>penicillin amidase</i>	Biosynthese von beta-lactam Antibiotika [359]
EC 3.6.3.25	<i>sulfate-transporting ATPase</i>	Transport von Sulfaten und Thiosulfaten [360]
Archaea:		
Enzym	Name	Pathway-Ansatz
EC 1.1.1.23	<i>histidinol dehydrogenase</i>	Finaler Schritt in der Histidinbiosynthese [361]
EC 2.5.1.89	<i>tritrans, polycis-undecaprenyl-diphosphate Synthase</i>	Isoprenoid Biosynthese [362]
EC 2.6.1.51	<i>serine-pyruvate transaminase</i>	Serin-Metabolismus: Serin zu Alanin und Hydroxypyruvat [20]
EC 3.6.3.4/54	<i>Cu⁺-exporting ATPase</i>	Ag ⁺ /Cu ⁺ -abhängige ATPasen-Aktivität [363,364]
EC 3.6.3.30	<i>Fe³⁺-transporting ATPase</i>	Fe ³⁺ -abhängige ATPasen-Aktivität nur bekannt bei Eukaryota, Bacteria [365]
EC 4.2.1.20	<i>tryptophan synthase</i>	Tryptophan Biosynthese: Indol zu L-Tryptophan [366]
EC 4.2.1.33	<i>3-Isopropylmalate dehydratase</i>	Initiierende Schritte vom Oxosuccinate zum Isopropylmalate in der Leucinbiosynthese [367]

Enzyme von Stoffwechselwegen weiterführender Datenbanken

Für einige Enzyme ist in anderen Datenbanken als MetaCyc schon eine Zuordnung im Sinne von alternativen Enzymen für einzelne Reaktionen zu finden. Die NAD(P)⁺-abhängige *glucose 1-dehydrogenase* (EC 1.1.1.47) wird sehr häufig bei den Rhodobacteraceen annotiert, ist jedoch nach vorliegendem Stand der Datenbank in keinem Pathway abgebildet. Für den ersten Schritt des Entner Doudoroff Weges, der direkten Umsetzung von Glucose zu einem Lacton, stellt dieses Enzym allerdings eine alternative Funktion dar [20] (Kapitel 3.4: Sauerstoffmetabolismus bei Archaeen). Ebenso kann die *serine-pyruvate transaminase* (EC 2.6.1.51) bei den Archaeen einem Weg des Serin-Metabolismus zugeordnet werden. So ist vor allem bei Eukaryoten die Katalyse der Verstoffwechslung von L-Serin und Pyruvat zu 3-Hydroxypyruvat und L-Alanin als Pyruvatzwischenpeicher nachgewiesen. [368] Des Weiteren ist eine Aktivität der *tryptophan synthase* (EC 4.2.1.20) dem Tryptophanmetabolismus zuzuordnen. In diesem Zusammenhang kann Tryptophan alternativ über Indol synthetisiert werden und dem Aufbau diverser Proteine zur Verfügung stehen. [366] Andere Enzyme, auffällig in den Rhodobacteraceen, spielen hauptsächlich in Transport- (*sulfate-transporting ATPase* (EC 3.6.3.25) [360]), Signal- (*histidine kinase* (EC 2.7.13.3) [357,358]) und Reaktionswegen als Antwort auf physiologischen Stress (*peroxiredoxin* (EC 1.11.1.15) [355], [*protein-Pil*] *uridylyltransferase* (EC 2.7.7.59) [356]) eine tragende Rolle. Sie können keinem Stoffwechselweg des Metabolismus im Sinne der Energiegewinnung und -umsetzung zugeordnet werden. Dies betrifft auch verschiedene katalytische Funktionen einiger metallionenabhängiger Transport-ATPasen bei den Archaeen (*Cu⁺-exporting ATPase* (EC 3.6.3.4/54) [363,364]). Einzelne können jedoch auch bisher nur bei Bakterien und Eukaryoten beschrieben werden (*Fe³⁺-transporting ATPase* (EC 3.6.3.30) [365]). Die *histidinol dehydrogenase* (EC 1.1.1.23) fungiert, ebenso nur bei Bakterien und Eukaryoten nachgewiesen, im finalen Schritt der Histidin-Biosynthese. [361] Allerdings kann auch gezeigt werden, dass dieser Syntheseweg sehr wahrscheinlich schon vor der evolutiven Aufspaltung der Domänen Archaea, Bacteria und Eukaryota entstanden ist. Infolgedessen können Genduplikationen, Genfusionen und zugleich auch der horizontale Gentransfer zwischen Archaeen und Bakterien vielseitige und individuelle Abwandlungen in Funktion und Genorganisation (Operons) hervorgebracht haben, während ein gewisser Grad der Sequenzhomologie verbleibt. [352] Die Prenylsynthase (EC 2.5.1.89) ist in die Isoprenoid-Biosynthese für einen Zucker-Transporter der Archaeenmembran involviert und bisher nur bei *Sulfolobus acidocaldarius* nachgewiesen. [362] In den meisten Datenbanken werden alle Prenylsynthasen (EC 2.5.1.81 – EC 2.5.1.91) noch keinem expliziten Pathway zugeordnet. Die Polyprenole sind allerdings unverzichtbare Elemente einiger Transporter und Transportwege für Zuckerbausteine. [369] Die meisten Archaeen nutzen hierfür im gleichen Sinne allerdings Dolichol statt Polyprenol. [370,371] Hierbei sind jedoch zum Beispiel die Enzyme *dolichyl-phosphate β-D-mannosyltransferase* (EC 2.4.1.83) oder *dolichol phosphate-mannose mannosyltransferase* (EC 2.4.1.B58) involviert. Diese sind in den vorliegenden Daten dementsprechend bei etwas über 50% aller untersuchten Archaeen annotiert. Beiden gemeinsam ist die Verarbeitung eines Monophosphates [372,373] anstelle des Diphosphates bei EC 2.5.1.89.

Schlussfolgerungen

Insgesamt zeigt das angewendete Verfahren einerseits, dass die gewählte Stoffwechselwegdatenbank MetaCyc, auf dem verwendeten Stand der Informationen, gerade auf alternative Zweige für verschiedene Reaktionsschritte der jeweiligen Pathways noch nicht verweist. Diesbezüglich ist eine Schnittmenge der Zuordnungen von MetaCyc, BRENDA und eventuell KEGG, insofern umsetzbar, für weitere Analysen empfehlenswert. Andererseits wird deutlich, dass Informationen und Gruppierungen zu Regulations-, Signal- und DNA-/RNA-Modifikationswegen eine weitere denkbare Möglichkeit zur Differenzierung phylogenetischer Gruppen und letztendlich auch deren Zusammenhänge mit den jeweiligen Metadaten darstellen können. Vorstellbare Quellen können zum Beispiel Datenbanken wie **SignalLink 2** [374] für Signalnetzwerke und **MODOMICS** [375] für klassifizierte RNA Modifizierungen sein. Auch Referenzdatenbanken zu einzelnen Enzymklassen, wie **MEROPS** [376] es beispielsweise für Peptidasen zur Verfügung stellt, können einbezogen werden. Eine weiterführende Recherche im Hinblick auf die Abdeckung der verschiedenen phylogenetischen Domänen ist in naher Zukunft anzustreben.

3.9 Das *Core-/Pan-„Enzymom“*

Im nachstehenden Kapitel werden die Gesamtheit aller annotierten Enzymklassen der analysierten Organismusgruppen (genannt *Pan-„Enzymom“*), sowie nur jene Funktionen, die stark vertreten sind (genannt *Core-„Enzymom“*), im Hinblick auf die Diversität und Vergleichbarkeit innerhalb phylogenetischer Kladen diskutiert (Kapitel 2.9: Das Vorgehen nach Verfahren II).

Allgemeine Fakten

Das *Core-„Enzymom“* der Archaeen ist mit 77 Enzymen relativ zum „Enzymom“ der Rhodobacteraceen mit 430 Enzymen recht klein. Allerdings ist hierbei nicht zu vergessen, dass es sich bei den Archaeen um eine ganze Organismendomäne handelt. Dennoch ist festzuhalten, dass im Median alle einbezogenen Archaeen 502 eindeutige EC Nummern im Genom codieren (abzüglich von Duplikaten). Das entspricht ebenfalls im Median 22% des Genoms bei einer medialen Genomgröße von 2277 Genen. Die Rhodobacteraceen weisen im Vergleich das größere mediale Genom der untersuchten Bakterien von 4145 Genen auf. Im Median codieren 851 Gene, bzw. 21% des Genoms für eine enzymatische Funktion. Die kleinste enzymatische Vielfalt ist mit 11% des Genoms bei *Sulfolobus solfataricus* P2 festzustellen. Das Pendant der Rhodobacteraceen ist *Labrenzia* sp. DG1229 mit verzeichneten 13%. Die größte Menge an einmalig gezählten Enzymen tritt bei dem Archaeon *Candidatus nitrosopelagicus* V2 mit 30% des Genoms auf. 28% des Genoms von *Loktanella vestfoldensis* SKA53 stehen dem aus der Gruppe der Rhodobacteraceen gegenüber.

Die Tabelle 19 zeigt, wie die Anzahl der *Core*-Enzyme der jeweiligen Organismengruppe zunimmt, je geringer der Mindestanteil an entsprechend annotierten Organismen ist. Diese Zunahme ist zu erwarten. Vergleicht man die „Enzymome“ (-95%) und (-70%), lassen sich allerdings Unterschiede bezüglich der Dimension des Anstiegs zwischen den Archaeen und den Rhodobacteraceen feststellen. Während sich das „Enzymom“ der Archaeen mehr als verdreifacht, erhöht sich die Anzahl der gemeinsamen Enzyme der Rhodobacteraceen lediglich auf das 1,6-fache. Das somit weitaus stabilere *Core-„Enzymom“* der Rhodobacteraceen zeigt, dass in dieser phylogenetischen Gruppe doch auf sehr ähnliche Strategien des Gesamtmetabolismus gesetzt wird. Dahingegen wird unterstützt, dass Archaeen, bzw. definierte Kladen mit deren Anpassungen an verschiedenste Extrema natürlich auch diverse Anpassungen ihres Stoffwechsels aufweisen müssen. [377]

Dem *Core-„Enzymom“* gegenüber steht das *Pan-„Enzymom“*, welches beschreibt, wie vielfältig der Enzympool der Organismengruppen ist. Der zuvor angesprochene Unterschied fällt hier mit etwa 1900 einfach gezählten Enzymen bei beiden Gruppen nicht auf. Reduziert man das *Pan-„Enzymom“* auf Annotationen, die mindestens in mehr als 10 verschiedenen Organismen auftreten, reduziert sich in beiden Fällen die Menge an enzymatischen Funktionen auf etwa das 0,6-fache.

Zusammengefasst lässt sich sagen: Die enzymatische Diversität der Domäne der Archaeen und der Familie der Rhodobacteraceen insgesamt ist vergleichbar groß. Hingegen ist bei den Vertretern der Bakterien einer taxonomischen Familie auf Grund redundanter Stoffwechselstrategien im homogeneren Habitat auch ein entsprechender gemeinsamer Anteil des Metabolsimus zu erwarten. Die Varianz an enzymatischen Funktionen und damit der Anpassung des Metabolismus an sehr verschiedene Anforderungen durch die Umwelt- und Lebensbedingungen spiegelt das relativ gesehen kleine Core-„Enzymom“ der Archaeen wieder.

Die Core-„Enzyme“

Zum erwähnten kleinen Kern der Enzyme in allen untersuchten Archaeen zählen gehäuft grundlegende Enzyme im Zusammenhang mit der Replikation (Polymerasen, Purin Metabolismus: z.B. EC 2.7.6.1, EC 2.7.7.7, EC 2.7.7.6), und Transferasen zur Modifikation von RNA (z.B. EC 2.7.7.72), sowie entsprechende Ligasen (EC 6.1.1.2, EC 6.1.1.3, EC 6.1.1.7, EC 6.1.1.10). Aber auch Enzyme im Zusammenhang mit der Biosynthese von Membranlipiden zur Bildung besonderer Membranstrukturen, um bei extremen Umweltbedingungen existieren zu können, gehören zum Core-„Enzymom“ der Archaeen. Dazu zählt zum Beispiel die Etherbindung zum Glycerat von gesättigten Polyprenolen und Isoprenoiden anstelle der Veresterung mit Fettsäuren (EC 2.5.1.29, EC 2.5.1.41, EC 2.5.1.89) [20]. Dazu fallen Enzyme des Mevalonat Stoffwechselweges wie die *mevalonate kinase* (EC 2.7.1.36) [20] im Core-„Enzymom“ auf. Auch dies ist am Aufbau von Isoprenoiden beteiligt. [20] Der Einbau dieser längerkettigen Kohlenstoffverbindung mit einer Etherbindung zum Glycerat unterscheidet die Archaeen von den meisten Bakterien und erhöht die Schutzleistung der Membran unter Extremstbedingungen. [282]

Die vergleichsweise breitere Verankerung enzymatischer Funktionen im Enzympool der Rhodobacteraceen zeigt neben den oben schon erwähnten grundlegenden Aufgaben in der DNA-/RNA-Polymerisierung und -Modifikation, bzw. dem Katabolismus und Anabolismus verschiedener Aminosäuren (Glutamin-Metabolismus: EC 1.4.1.13, Asparaginsäure-Metabolismus: EC 2.6.1.1, Threonin-Metabolismus: EC 4.3.1.19) auch weitere Hinweise auf Kladenumfassende Annotationen. So scheinen fest verankerte Reaktionsmechanismen in der Klade dazu zu dienen, auf oxidativen Stress zu reagieren, bzw. Sauerstoffradikalen entgegenzuwirken (EC 1.15.1.1) [20]. Auch benötigte Enzyme der Biotin (Vitamin B7) und Vitamin B12 Biosynthese sind offensichtlich bei den Rhodobacteraceen verbreitet (EC 6.3.4.14, EC 6.3.4.15, EC 6.6.1.2) [20]. Dies ist auf die Biofilmbildung zusammen mit Biotin/Vitamin B12 auxotrophen Algen [378,379] im hauptsächlich marinen Lebensraum zurückzuführen. [65]

Die Pan-„Enzynom“

Schaut man auf Enzyme, die Reaktionen in Bezug auf die Verstoffwechselung bestimmter Metabolite katalysieren, so lassen sich durchaus verschieden entwickelte Wege vermuten (Pan-„Enzynom“). So ist die Aufnahme von Stickstoff zum Beispiel nur bei einer Teilgruppe der Rhodobacteraceen mit der Nitrogenase EC 1.18.6.1 und Ferredoxin als Elektronendonator unterstützt (vorrangig einige *Rhodobacter* Spezies und explizit Stämme von *Rhodobacter capsulatus*). Die Denitrifikation hingegen ist einer anderen Gruppe, primär der Paracoccen zugesprochen (EC 1.7.2.4, „Nitrous-oxide reductase“) [20]. Die Energiegewinnung durch die Methanogenese zeigt sich in der Domäne der Archaeen in mehreren Varianten kleinerer Untergruppen ebenfalls in verschiedenen Stoffwechselwegen. So sind vor allem die *Methanosarcina* Spezies mit den Enzymen EC 2.1.1.251 und der Kombination aus EC 2.1.1.249/247 dazu fähig, Dimethylsulfide, bzw. Dimethylamine als Ausgangsstoff zu verwerten. Für viele andere methanogene Bakterien hingegen sind Wasserstoff H_2 und Kohlenstoffdioxid CO_2 die Ausgangsstoffe für einen methanbildenden Energiegewinnungsweg (EC 1.2.99.5, EC 2.3.1.101, EC 3.5.4.27, EC 1.12.98.2, EC 1.5.98.2, EC 2.1.1.86). [21]

Tabelle 19: Core-/Pan-„Enzynom“ der Projektgruppen. Das Core- und das Pan-„Enzynom“ der beiden Organismengruppen werden in dieser Tabelle gegenübergestellt. Dabei wird die eindeutige Anzahl der annotierten Enzyme mit Bezug auf den prozentualen Anteil der vertretenen Organismen zur Gesamtgruppe aufgelistet. Dementsprechend ist die Größe des Core-„Enzynoms“ der jeweiligen Gruppe mit einer 95-, 90-, 80- und 70-prozentigen Abdeckung der untersuchten Gruppe zu sehen. Bei dem Pan-„Enzynom“ kann unterschieden werden, ob Enzyme mindestens in einem, mehr als einem oder mehr als zehn verschiedenen Organismen annotiert sein müssen, um zur Menge des Enzympools der jeweiligen Gruppe gezählt zu werden.

Core-„Enzynom“	Projektgruppe	
	n Rhodobacteraceae	n Archaea
95% der Projektgruppe	430	77
90% der Projektgruppe	527	115
80% der Projektgruppe	636	191
70% der Projektgruppe	688	263
Pan-„Enzynom“	Projektgruppe	
	n Rhodobacteraceae	n Archaea
alle Organismen	1891	1907
mehr als ein Organismus	1711	1683
mehr als zehn Organismen	1262	1208

Schlussfolgerungen

Die Vielseitigkeit der Archaeen [377] und der breit aufgestellte Metabolismus der erfolgreichsten und häufigsten Gruppe im marinen Habitat, den Rhodobacteraceen [78], machen die Problematik des Quervergleichs von Organismen einer Klade deutlich. Gefestigte Anpassungen an das jeweilige Habitat können sich im charakteristischen Auftreten von Enzymen widerspiegeln (Core-„Enzymom“), jedoch können auch verschiedene Strategien in der Evolution der Klade entstanden sein (Pan-„Enzymom“). Die Annahme von Enzymfunktionen ohne direkten sequenzhomologen oder experimentellen Nachweis für einen expliziten Organismus durch den Vergleich zu Schwesterorganismen der Klade, kann demzufolge offensichtlich zu Fehlinterpretationen führen. Verdeutlicht werden diese Differenzen bei der Analyse der phylogenetischen Verwandtschaftsverhältnisse auf der Grundlage der annotierten Enzymklassen im nächsten Kapitel.

3.10 Die phylogenetischen Bäume auf Basis binärer Enzymvektoren

Die Betrachtungen in diesem Kapitel vergleichen die Darstellung der phylogenetischen Verwandtschaftsverhältnisse auf Grundlage der Sequenzen der 16S rRNA der jeweiligen Organismen der Analyseebene mit den resultierenden Distanzen der Vertreter basierend auf ihrem individuellen Enzympool (Kapitel 2.9: Das Vorgehen nach Verfahren II). Infolgedessen sollten, konform zum häufig in der Modellierung von Organismen angewendeten Quervergleich der Enzymannotationen als Grundlage für nicht experimentell, bzw. bioinformatisch nachgewiesene Funktionsvorhersagen (Kapitel 1.4), keine gravierenden Differenzen auftreten.

Die 16S rRNA-Analyse der Rhodobacteraceen

Der phylogenetische Baum der 16S rRNA-Analyse der Rhodobacteraceen in Abbildung 9 entspricht den allgemein anerkannten phylogenetischen Verwandtschaftsverhältnissen dieser Gruppe. [380,381] Die Außengruppe der Schwestertaxa *Stappia* und der bekannterweise eher mit den Rhizobiales zusammenfallenden Rhodobacteraceen *Ahrensia*, *Rhodovulum*, *Pannonibacter* und *Polymorphum* [74,381] tritt randständig in Erscheinung, so dass hier eine Wurzel mittels „Midpoint-Rooting“ (Kapitel 2.3 und Kapitel 2.9: Das Vorgehen nach Verfahren II) gezogen wird. Das „Midpoint-Rooting“-Verfahren setzt für eine Wurzel eines phylogenetischen Baumes die größte Distanz zwischen zwei beinhalteten Taxa oder Kladen voraus. Dies ist demzufolge mit der Distanz der Rhodobacteraceen zur bewusst gewählten Außengruppe der Fall. Die Vertreter der Spezies *Oceanicola* verteilen sich auffallend über das gesamte Cluster der Rhodobacteraceen. Daher wird eine Reklassifizierung dieser Stämme in einigen Publikationen schon angestrebt. [74,78,382,383] Die Differenzierung in ein marines und nicht marines Habitat zeigt keine Clusterbildung. Vielmehr sind nicht marine im Erdreich lebende Organismen (*Ketogulonicigenium*, *Rubellimicrobium*, *Paracoccus*) vorrangig in Gruppen mit dem Habitat Sediment eingegliedert. Ansonsten können in Bezug auf die Habitate keine nennenswerten Clusterbildungen erkannt werden.

3 Ergebnisse und Diskussion

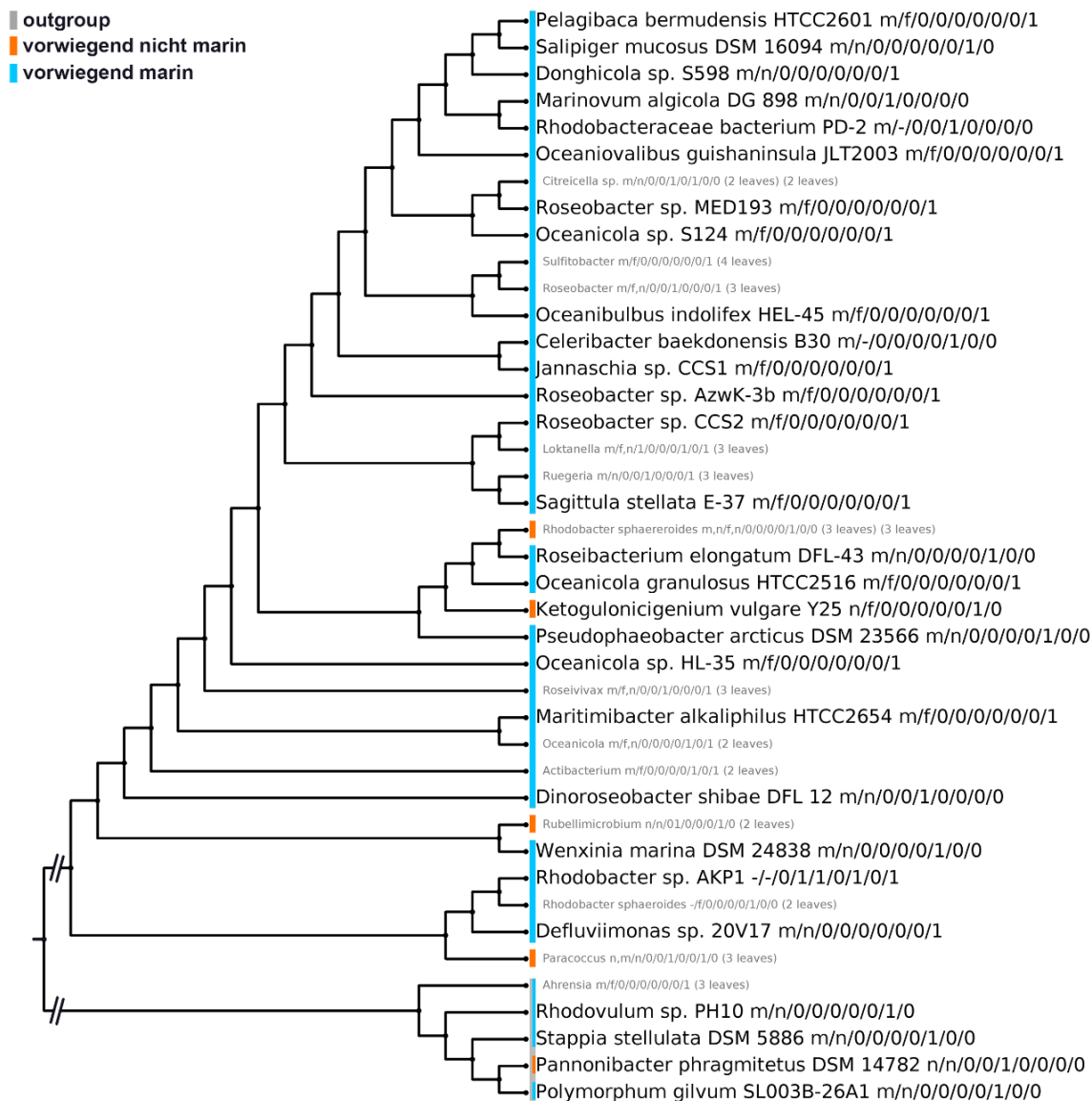


Abbildung 9: Neighbor Joining-Baum der 16S rRNA Sequenzen der Rhodobacteraceen. Als Outgroup dienen nahe Verwandte der Organismengruppe (Kapitel 2.9: Das Vorgehen nach Verfahren II).

Metadaten-Vektor (nach Spezies stehend):

marin (m) | nicht marin (n) | keine Angabe (-) /

freilebend (f) | nicht freilebend (n) | keine Angabe (-) /

Habitat: Biofilm (1|0) / Frischwasser (1|0) / Organismus-gebunden (1|0) / Meeres (1|0) / Sediment (1|0) /

Erdreich (1|0) / Wasser (1|0)

Binäre Enzymvektoren der Rhodobacteraceen

Der Distanzbaum auf Basis der Enzymvektoren (Abbildung 10) erhält nahezu die Außengruppe randständig. Lediglich scheint *Octadecabacter antarcticus* 307, ein psychrophiles, in Gasvakuolen des Polarmeeres lebendes Bakterium [384], eine geringere Distanz zum positiv-Außengruppenvektor zu haben. Der in diesem Verfahren verwendete positiv-Außengruppenvektor, bildet einen Vektor ab, der alle annotierten Enzymklassen der untersuchten Gruppe von Organismen scheinbar bedient (Kapitel 2.9: Das Vorgehen nach Verfahren II). Der oben angesprochene *Octadecabacter*-Stamm hat demzufolge einen entsprechend großen Enzympool zur Verfügung und ist für weitergehende Forschungen psychrophiler Organismen [384] interessant. Ebenso clustern auch hier die *Oceanicola* Stämme nicht zusammen. Dass diese Vertreter sich also auch in der Annotation von Enzymen und damit dem Metabolismus unterscheiden, unterstreicht die Forderung nach einer Reklassifizierung. Nicht marine Organismen bilden auch hier keine monophyletische Gruppe. Ganz im Gegensatz dazu gliedern sie sich teilweise eng in marin, wasserlebende Gruppierungen ein (*Paracoccus/Actibacterium mucosum*, *Paracoccus denitrificans/Ruegeria pomeroyi*, *Rhodobacter capsulatus/ Actibacterium* sp.). Das *Actibacterium* ist ein chemo-organotropher Stamm, der im Jahre 2012 erst als eigenständiger Genus klassifiziert wurde. [385] Der Zusammenhang hier bedarf einer weiteren Untersuchung und ist offensichtlich nicht in Verbindung mit den bekannten Metadaten zu sehen. Des Weiteren ist eine Gruppe organismen-assoziierteter Bakterien, bestehend aus *Roseibium*, *Labrenzia*, *Pannonibacter* und *Pseudovibrio*, zu nennen. Hier können clustertypische Enzyme zum Beispiel zur Biofilmbildung charakterisierend sein. [386] Abschließend bleibt zu erwähnen, dass im enzymatischen Ansatz eng verwandte Spezies-Stämme nicht immer als Schwestergruppen erhalten bleiben. So verteilen sich zum Beispiel die *Phaeobacter* über den gesamten Baum. Dies lässt auf einen sehr unterschiedlichen Pool von Enzymen und damit Differenzen im Metabolismus des Genus schließen.

3 Ergebnisse und Diskussion

- outgroup
- vorwiegend nicht marin
- vorwiegend marin

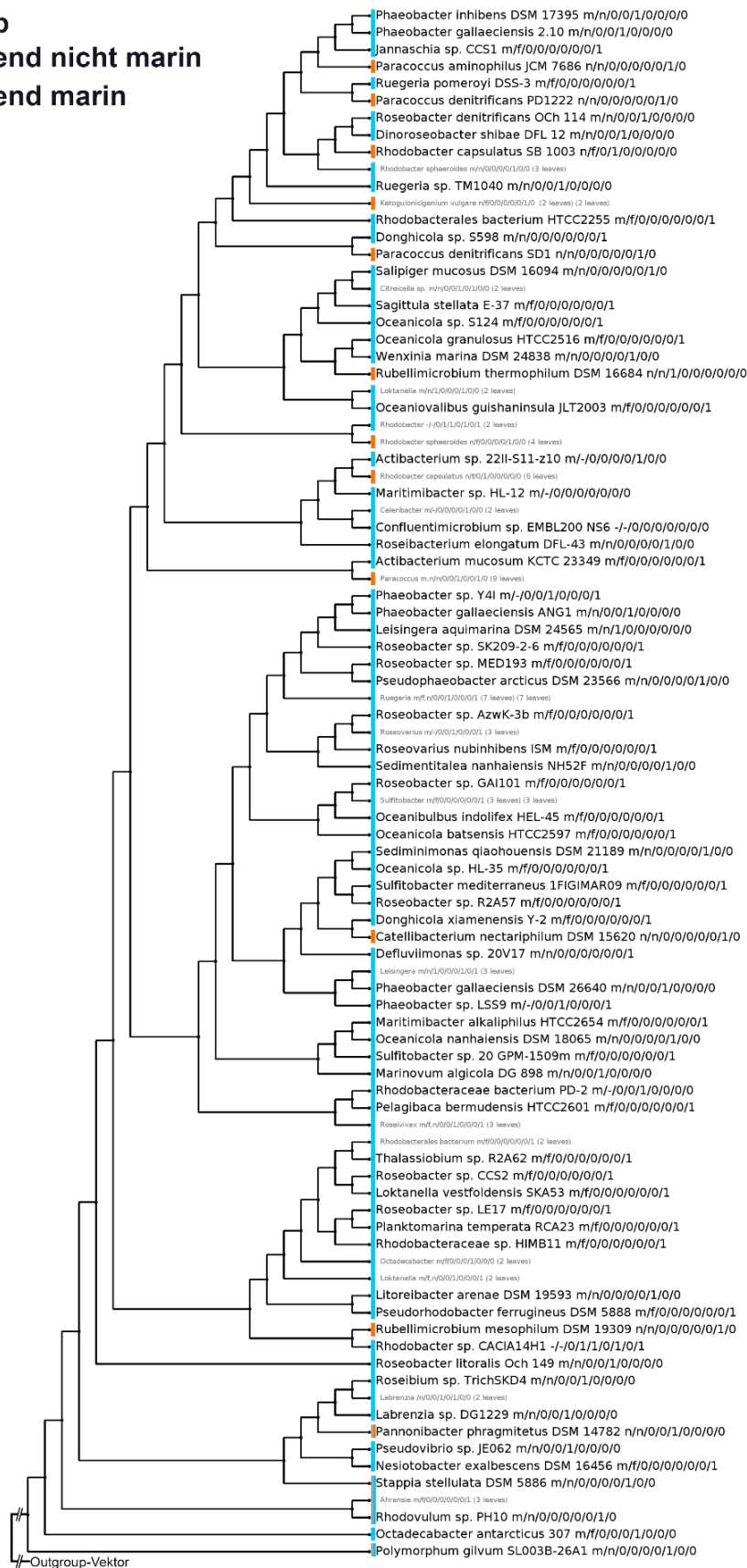
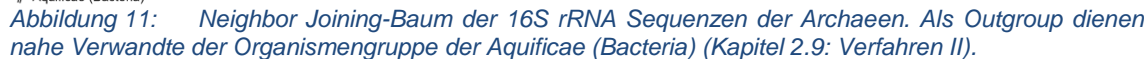


Abbildung 10: Neighbor Joining-Baum der binären Enzym-Vektoren der Rhodobacteraceen. Als Outgroup dient ein Positivvektor aller Enzyme der untersuchten Organismengruppe (Kapitel 2.9: Das Vorgehen nach Verfahren II). **Metadaten-Vektor (nach Spezies stehend):** siehe Abbildung 9.

Die 16S rRNA-Analyse der Archaeen

Die Übergruppierungen in Euarchaeota (halophil und methanogen), Crenarchaeota (thermophil) und Kor-/Thaumarchaeota der 16S rRNA Analyse in Abbildung 11 entspricht der allgemein gültigen phylogenetischen Auffassung. [377,381] Eine Wurzel des phylogenetischen Baumes kann mit den Aquificae, einer Klasse der Bakterien, die jedoch in großen Teilen des Genoms sehr ähnlich zu den Archaeen ist [278,387,388], als Außengruppe eingezeichnet werden („Midpoint-Rooting). Eine Sondergruppe der Klasse der Thermoplasmata ist durch die Spezies *Thermogymnomonas*, *Picrophilus*, *Ferroplasma*, *Thermoplasma* und *Aciduliprofundum* vertreten. Dabei handelt es sich um Euarchaeota, die eine Sonderstellung auf Grund ihrer extrem sauren, bzw. heißen Habitate einnehmen (acidophil, thermophil). [389,390] Vorrangig aerob klassifizierte Spezies sind die Vertreter der Sulfolobales und Metallosphaera im thermophilen Bereich, *Picrophilus* der Thermoplasmata und vor allem halophile Eukaryoten (*Natrialba*, *Haloterrigena*, *Halopiger*, *Natronococcus*, *Salinarchaeum*, *Halogeometricum*, *Halorubrum*, *Halomicrobium* und *Haloarcula*). Methanogene Archaeen bilden wie erwartet eigenständige Subcluster. [377] Verschiedene Stämme gleicher Spezies lassen sich durch Ähnlichkeit der 16S rRNA und somit der direkt abzubildenden Verwandtschaft als Cluster zusammenfassen (z.B. Sulfolobales, Candidatus, *Methanosarcina*, *Halorubrum*). Festzuhalten sind abschließend die Eigenschaften der Halophilie, Thermophilie und der Methanogenese als oberste Ebene der Clusterbildung. Metadaten der Acidophilie/Acidophobie, des Typus des Metabolismus und des Habitats ordnen sich dieser Ebene unter und führen hier allerdings zu keinen klaren Abgrenzungen von Kladen.



pH-Optimum: von: <6,5(pHf-s) | 6,5-7,5(pHf-n) | >7,5(pHf-a) / bis: <6,5(pHt-s) | 6,5-7,5(pHt-n) | >7,5(pHt-a)

Sauerstoffmetabolismus: anaerob | aerob | fakultativ /

Schmutzwasser (0|1) /

Metabolismus: chemotroph (0|1) / litotroph (0|1) / organotroph (0|1) / autotroph (0|1) / heterotroph (0|1) /

Binäre Enzymvektoren der Archaeen

Die Enzymvektoren und die darauf basierenden Distanzen der Archaeen sind in Abbildung 12 dargestellt. Die grundlegende Gruppierung nach halophil, thermophil und methanogen infolge der Abteilungen Crenarchaeota, Euarchaeota und Korarchaeota bleibt auch hier erhalten. Die Sondergruppen der Kor-/Thaumarchaeota und der *Thermoplasmata* können hingegen auf Ebene der Enzyme nicht erhalten bleiben. Hingegen gliedern sie sich entweder vorrangig an chemo-, lito-, autotrophe Archaeen an (*Candidatus*, *Thermoplasmatales archaeon* → *Ferroglobus*, *Archaeoglobus*) oder sie fallen in ein Cluster mit vorrangig heterotrophen Thermophilen (*Candidatus korarchaeum*, *Aciduliprofundum* → *Hyperthermus butylicus*, *Thermophilum*, *Stapylothermus*). Weiterhin gruppieren sich die extremen Euarchaeota *Picrophilus* und *Thermoplasma volcanium* direkt mit zwei autotrophen, acidophilen Spezies der Sulfolobales zusammen. Darüber hinaus lässt sich feststellen, dass verschiedene Stämme, bzw. verschiedene Spezies eines Genus nicht mehr so zusammen clustern, wie es die Phylogenie scheinbar vorgibt. Am Beispiel der Sulfolobales ist zu erkennen, dass die Distanzgrundlage auf Enzymbasis dazu führt, zwei getrennte Gruppen zu erhalten. So sind auf Enzymebene die Spezies *Sulfolobus solfataricus* und *Sulfolobus acidocaldarius* als aquatisch lebende autotrophe Archaeen enger miteinander verwandt als die oftmals heterotroph deklarierten *Sulfolobus islandicus* und *Sulfolobus tokodaii*. Dies und der Zerfall der Klade der Thermoplasmata machen deutlich, dass hier die Metadaten des Organismus, wie die Art des Metabolismus und das Habitat mit zum Beispiel den pH-Eigenschaften, eine entscheidendere Rolle als auf rein phylogenetischer Grundlage spielen können. Hingegen gibt es kladencharakterisierende Eigenschaften wie die Methanogenese, die Thermophilie und die hohe Halotoleranz, die die phylogenetische Sichtweise unterstützen, während sich viele definierte Habitate und der Sauerstoffmetabolismus nicht in einer Änderung der Großgruppen niederschlagen. Insgesamt ist noch zu bemerken, dass die Euarchaeota, speziell die extrem halophilen Spezies, die kürzeren Distanzen zum positiven Outgroupvektor (Kapitel 2.9: Das Vorgehen nach Verfahren II) aufweisen. Da bei diesem davon ausgegangen wird, dass alle bekannten Enzyme vorhanden sind, besitzen die Euarchaeota also tendenziell mehr dieser Enzymmenge als die Crenarchaeota im thermophilen Bereich. Die demzufolge mutmaßlich mächtigere Genomgröße und der größere Enzympool der halophilen Archaeen spiegeln die beträchtlichere Vielfältigkeit des Metabolismus von Organismen wieder, die verschiedenste Zucker als Kohlenstoffquellen verwenden und im Labor auch auf Minimalmedien in jedem Fall unter extremsten Salzstress existieren können. [391]

3 Ergebnisse und Diskussion

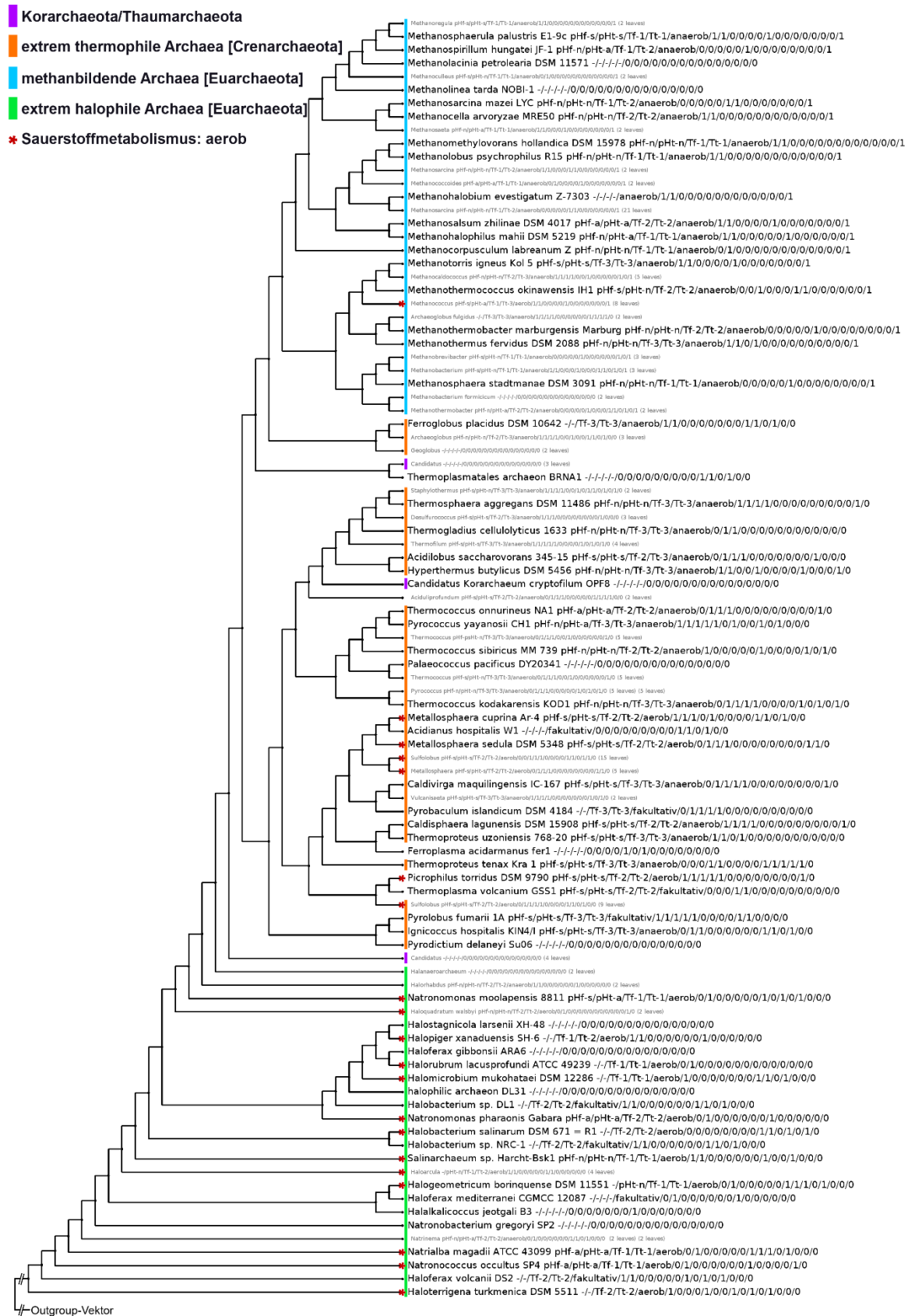


Abbildung 12: Neighbor Joining-Baum der binären Enzym-Vektoren der Archaeen. Als Outgroup dient ein Positivvektor aller Enzyme der untersuchten Organismengruppe (Kapitel 2.9: Das Vorgehen nach Verfahren II). Metadaten-Vektor (nach Spezies stehend): siehe Abbildung 11.

Schlussfolgerungen

Zusammenfassend ist zu sagen, dass die Phylogenie im Großen und Ganzen nur durch den Vergleich, bzw. die Distanz der Enzymvektoren der Organismen auf Phylum-Ebene unterstützt werden kann. In der Verwandtschaft von Spezies und verschiedenen Stämmen dieser können allerdings disgruente Schlussfolgerungen durch Differenzen im Metabolismus erfolgen. Die hier dargestellte Phylogenie wird damit nicht in Frage gestellt. Dies muss lediglich beim Betrachten von und Schlussfolgern aus den Metabolismus betreffenden Vergleichen berücksichtigt werden.

3.11 Die phylogenetischen Bäume auf Basis binärer Pathway-Vektoren

Phylogenetische Bäume lassen sich nicht nur, wie im vorherigen Kapitel dargestellt, auf der Basis von 16S rRNA-Sequenzen und Enzymvektoren erstellen. Die individuelle Zuordnung von Stoffwechselwegen (Kapitel 3.6) lässt ebenso die Berechnung von Distanzen auf Grund von Unterschieden der entsprechenden Binärvektoren der analysierten Organismen zu. Dies entspricht einem grundlegenden Vergleich der metabolischen Zusammenhänge und Anpassungen der Vertreter unterschiedlicher Habitats und gibt Anlass zur Diskussion, ob in diesem Kontext phylogenetische Beziehungen aufrecht erhalten werden oder nicht.

Binäre Stoffwechselwegvektoren der Rhodobacteraceen

Der erstellte Baum auf Basis der Stoffwechselwege der Rhodobacteraceen (Abbildung 13) lässt die aus der 16S rRNA Analyse bekannte Außengruppe nicht mehr als zusammenhängenden Block erkennen. Der phylogenetische Zusammenhang scheint hier also weniger fundamentiert als noch auf Enzymebene. Entscheidende Cluster, die sich konkret auf ein bestimmtes Habitat zurückführen lassen, sind leider nicht zu erkennen. Lediglich die meisten nicht marinen Organismen (*Paracoccus*, *Ketogulonicigenium*, *Rhodobacter capsulatus*, *Rhodobacter sphaeroides*) gruppieren sich deutlich zusammen mit den im Sediment oder Boden lebenden marinen Vertretern (*Roseibacterium*, *Rhodobacter* sp., *Rhodovulum*). Dies ist höchstwahrscheinlich auf typische Fermentationen und Nitratreduktionen in anaeroben Umgebungen zurückzuführen.

3 Ergebnisse und Diskussion

■ outgroup

■ vorwiegend nicht marin

■ vorwiegend marin

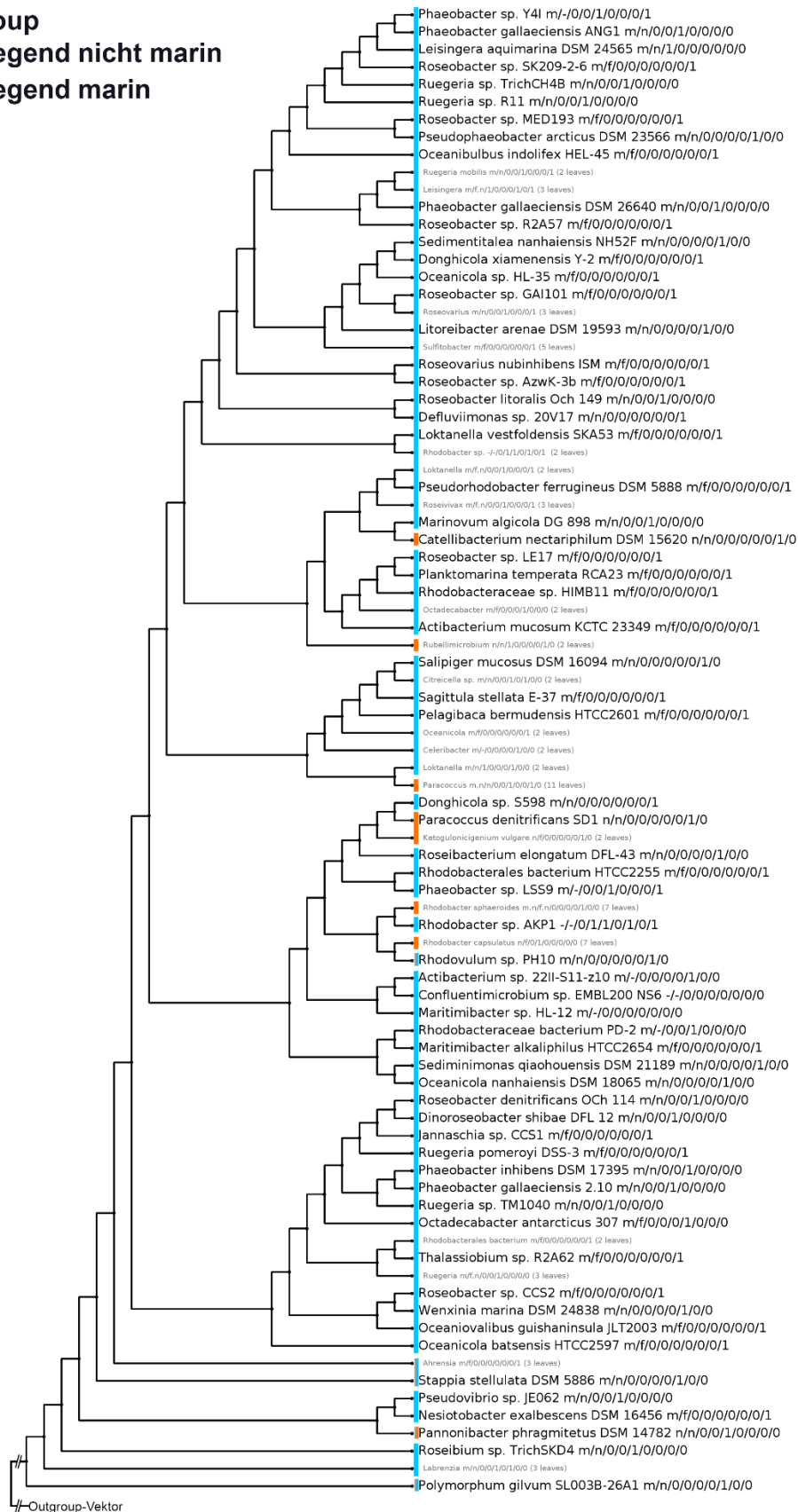


Abbildung 13: Neighbor Joining-Baum der binären Pathway-Vektoren der Rhodobacteraceen. Als Outgroup dient ein Positivvektor aller Pathways der untersuchten Organismengruppe (Kapitel 2.9: Das Vorgehen nach Verfahren II). Metadaten-Vektor (nach Spezies stehend): siehe Abbildung 9.

Binäre Stoffwechselwegvektoren der Archaeen

Ebenso wie der enzymbasierte Baum der Archaeen zeigen die Stoffwechselwege dieselben Hauptgruppierungen der Crenarchaeota und der Euarchaeota inklusive der methanogenen und halophilen Unterscheidungen (Abbildung 14). Die geringe Distanz der halophilen Archaeen zum Positivvektor (Kapitel 2.9: Das Vorgehen nach Verfahren II), der alle Pathways beinhaltet, lässt sich, wie im Kapitel 3.9 schon erwähnt, auf die Vielzahl der vorhandenen katabolischen Abbauewege verschiedener Zucker der salzstressresistenten Organismen zurückführen. [391]

Bis auf *Candidatus korarchaeum cryptofilum* gruppieren sich hier allerdings die Kor-/Thaumarchaeota wieder nahezu zusammen. Das bedeutet, dass viele der Enzyme, die im Binärbaum noch zu Differenzen dieser Klade führen, keinen Stoffwechselwegen zugeordnet werden können. Dies ist für weitergehende Forschungen von Interesse.

Die Aufspaltung der Thermoplasmata hingegen wird bestätigt. So gliedern sich die Vertreter der *Aciduliprofundum* an eine Gruppe der Thermo- und Pyrococcen, die ebenfalls oft ein pH-Optimum im alkalischen Bereich nachweisen. Dagegen nimmt das *Thermoplasmatale archaeon* BRNA1 eine Stellung in naher Verwandtschaft in der Kor-/Thaumarchaeota-Gruppe ein. Die über alle drei Bäume annähernd beständige Gruppe der übrigen Thermoplasmata *Ferroplasma*, *Thermoplasma* und *Picrophilus* gliedert sich auch hier nahe den acidophilen Vertretern der Crenarchaeota (Sulfolobales, *Metallosphaera*, *Thermoproteus*) ein.

In diesem Verfahren der Baumerstellung gruppieren sich die aeroben Crenarchaeota ebenso zusammen, wie im enzym-, bzw. 16S rRNA basierten Pendant. Allerdings ist auffällig, dass die Spezies *Sulfolobus solfataricus* ein zu den Crenarchaeota randständiges Cluster mit *Pyrococcus* und *Archaeoglobus fulgidus* einnimmt. Neben dem thermophilen Optimum beschreibt nur ein aquatisches, vulkanisches Habitat, bzw. heiße Quellen die gemeinsamen Metadaten. Ein zugeordneter Stoffwechselweg, den diese Gruppe gemeinsam hat, ist zum Beispiel der Aufbau von Spermidin („Spermidine Biosynthesis I“). Ausgehend von Spermidin werden zum Beispiel verzweigte Polyamine gebildet. Diese langkettigen und verzweigten Polyamine wiederum unterstützen die Stabilisierung von DNA und RNA. Somit kann höchstwahrscheinlich die Effizienz der Translation bei hohen Temperaturen gesteigert werden. [209]

Die Crenarchaeota-Gruppe der Archaeoglobi bestehend aus weiteren Spezies der Archaeoglobi, *Ferroglobus placidus* und *Geoglobus* ist im 16S rRNA und Enzymbaum randständig nahe den methanogenen Euarchaeota eingeordnet. Auf Ebene der Stoffwechselwege allerdings wird dieselbe Gruppe direkt in die Klade der methanbildenden Archaeen eingegliedert. Eine Begründung dafür kann eben in den Enzymen und Stoffwechselwegen im Zusammenhang mit der Methanbildung gefunden werden. Die Archaeoglobales sind dafür bekannt, ein RubisCO ähnliches Enzym in einem AMP Recycling Stoffwechselweg zu verwenden. [392] Weiterhin ist ein reduktiver Acetyl-CoA Pathway zur Kohlenstofffixierung nachgewiesen. [393] Dieser Stoffwechselweg weist einen Zweig zur Reduzierung des CO₂ zu einer Methylgruppe, ganz ähnlich zur

3 Ergebnisse und Diskussion

Methanogenese, auf. Ein weiterer Zweig bildet mit Hilfe eines zweiten CO_2 Moleküls die Carbonylgruppe, die dann zusammengefügt das Acetyl-CoA für den Kohlenhydratstoffwechsel bilden. [394]

Schlussfolgerungen

Festzuhalten bleibt, dass auch die Ebene der Stoffwechselwege die Großgruppierung der Phylogenie der Eu-, Cren, und Kor-/Thaumarchaeota bestätigt. Dazu können vor allem die Metadaten bezüglich der Halophilie, Thermophilie und Methanogenese herangezogen werden. Der Einfluss des Sauerstoffmetabolismus, des Metabolismus an sich und des Habitats kann nur in untergeordneten Gruppen wahrgenommen werden. Allerdings kann dieses Verfahren basierend auf der Distanz von Pathway-Vektoren der Phylogenie im Hinblick auf die Zusammengehörigkeit diverser Spezies und Stämme auch wie gezeigt widersprechen.

Schlussendlich ist festzustellen, dass der Metabolismus demgemäß direkter im Zusammenhang mit den Metadaten der Organismen steht, als es die phylogenetische Taxonomie teilweise vorsieht. Selbst bei direkten Schwestertaxa ist die Wahl der zu vergleichenden Organismen, um methodisch unvorhergesagte Enzymfunktionen in einem zu analysierenden Organismus vorauszusetzen, folglich definitiv nicht allein von den phylogenetischen Verwandtschaftsverhältnissen abhängig zu machen. Vielmehr sind die Anpassungen auf enzymatischer und infolgedessen metabolischer Ebene der Individuen zu bestimmen, einzuordnen und in Zusammenhang zu bringen, um fundamentierte Aussagen zu einzelnen zu annotierenden Funktionen treffen zu können.

4 Fazit und Ausblick

Das weiterentwickelte Programm namens EnzymeDetector schafft die Möglichkeit mit Hilfe von einem umfassenden und rationalen Bewertungsschema für die Enzymannotationen verschiedenster Bakterien und Archaeen in der Bioinformatik etablierter Annotationsdatenbanken und eigener Algorithmen, Konfidenzen für die wahrscheinlichsten enzymatischen Funktionen der jeweiligen Gene zu erkennen. Dies gestattet fundamentierte Interpretationen des Pools an Enzymklassen, die im Metabolismus eines Organismus und ferner ganzer Organismengruppen tragende Rollen einnehmen und in Zusammenhang mit dem Habitat und weiteren Umweltbedingungen zu stellen sind (Kapitel 3.5 und Kapitel 3.9). Durch die für diese Methode essentielle Unterscheidung obligatorischer Schlüsselfunktionen und alternativer Reaktionsketten in Stoffwechselwegen (Kapitel 3.2) werden diese Zuordnungen im Kontext stehend interpretiert und sowohl metadatentypische Anpassungen des Metabolismus der Organismen, als auch lebensnotwendige Mechanismen erkannt (Kapitel 3.6). In bisherigen Arbeiten wurde das hier gezeigte Verfahren entweder auf vereinzelte Organismen und Kleingruppen angewendet, oder nur explizite Stoffwechselfunktionen untersucht. Dementsprechend sind typische Abbauwege von Aminosäuren [79] und der Katabolismus verschiedener Kohlenhydrate [80] bei der Gattung der *Phaeobacter* aus der Familie der Rhodobacteraceae ebenso untersucht, wie die Synthese marintypischer Substrate im Umgang mit Salzstress und der Limitierung von Stickstoffquellen bei im Meer lebenden Vertretern der Bakterien, bzw. am Beispiel von *Dinoroseobacter shibae* und seiner phylogenetischen Familie [81]. Die vorliegende Arbeit verweist dahingegen auf domänentypische Zusammenhänge des gesamten Metabolismus der Vertreter der Archaeen und der Familie der Rhodobacteraceae mit den vorherrschenden Habitatbedingungen und stellt typische Verknüpfungen von Metadaten heraus. Demgemäß können sowohl die eingangs aufgestellten Leitthesen im Folgenden beantwortet und gefestigt, als auch der Hinweis auf die bedachte phylogenetische Anwendung auf dem Gebiet der Organismus-Modellierung bestätigt werden.

Die Beantwortung der Leitthesen

1. Die Annotation von Enzymen zu einem untersuchten Organismus ist ohne einen sequenzhomologen oder experimentellen Hinweis nur bedingt durch den Vergleich mit nahe verwandten Organismen möglich. Dies wird bestätigt durch die teils über die Erwartungen hinausgehende Vielseitigkeit des Metabolismus von Kladen und phylogenetischer Gruppen, bzw. Domänen. Infolgedessen äußern sich gefestigte Anpassungen an das jeweilige Habitat im charakteristischen Auftreten von Enzymen (Core-„Enzymom“, Kapitel 3.9), jedoch können auch verschiedene Strategien in der Evolution einer Klade entstanden sein (Pan-„Enzymom“, Kapitel 3.9), um auf Umwelteinflüsse zu reagieren und sich anzupassen. Die Annahme von Enzymfunktionen ohne direkten sequenzhomologen oder experimentellen Nachweis für einen expliziten Organismus durch den Vergleich zu Schwesterorganismen der Klade, Familie oder innerhalb der Domäne, kann demzufolge offensichtlich zu Fehlinterpretationen führen. Der bei der Erstellung metabolischer Modelle von Organismen oft und weit verbreitete Entschluss, nicht eindeutig verifizierte Enzymfunktionen, Metabolite und Mechanismen mit Hilfe gut untersuchter Feststellungen im Quervergleich zu anderen Organismen zu beschreiben [29], führt folglich in vielen Fällen unweigerlich zu einem Trugschluss. Wie weiterhin festgestellt folgt der Metabolismus nicht ausschließlich phylogenetischen Verwandtschaftsbeziehungen, sondern spiegelt immer die Beziehungen und Anpassung der Organismen an ihre Umwelt wieder. (Kapitel 3.10 und Kapitel 3.11)
2. Die Untersuchung phylogenetischer Kladen zeigt auf Enzym-, bzw. Stoffwechselwegebene Unterschiede zum allgemein anerkannten 16S rRNA Verfahren. Auf Phylumebene, bzw. der Einteilung von Großgruppen und phylogenetischen Abteilungen sind kaum Widersprüche zu definieren. Dementsprechend bleibt bei den Archaeen beispielsweise die Einteilung der Abteilungen der Crenarchaeota, Euarchaeota und Kor-/Thaumarchaeota weitestgehend bei der Untersuchung auf Basis der Distanzen von Enzym pools (Kapitel 3.10) und Stoffwechselwegen (Kapitel 3.11) bestehen. Großgruppentypische Eigenschaften des Metabolismus wie zum Beispiel die Methanogenese oder die Halotoleranz bestätigen die phylogenetische Zusammengehörigkeit deutlich. Im Sinne von engeren Verwandtschaften wie Schwestertaxa oder verschiedener Spezies einer Art zeigen sich hingegen Differenzen. Ein entsprechendes Beispiel stellt die Sondergruppe der temperatur- und pH-extremen Thermoplasmata der Euarchaeota dar. Die Phylogenie gruppiert sie zusammen, während die Auswertung der enzymatischen Funktionen und der zugeordneten Stoffwechselwege eine deutliche Unterscheidung auf Grundlage der Acidophilie, bzw. Alkaliphilie darstellt. Ebenso gliedert sich die Ordnung der Sulfolobales nur auf metabolischer Grundlage auf in aquatisch lebende Vertreter, die Mechanismen zur Fixierung von Kohlenstoffdioxid aufweisen (*Sulfolobus solfataricus* und *Sulfolobus acidocaldarius*), und in entsprechender Distanz die oftmals heterotroph deklarierten *Sulfolobus*

islandicus und *Sulfolobus tokodaii*, die rein organische Substrate verstoffwechseln.

Zusammengefasst wird folglich wie in Punkt 1 deutlich: Sollen phylogenetische Quervergleiche der weiteren Beschreibung des Metabolismus eines Organismus dienen, so ist nicht nur die reine Verwandtschaft zu berücksichtigen, sondern vielmehr sind die entsprechenden Metadaten und Informationen zu den Habitaten der Vertreter zu sammeln und zu vergleichen. Denn ihr Einfluss ist wie nachgewiesen von hoher Bedeutung und führt unweigerlich zu Differenzen in Bezug auf zu annotierende Enzyme und Stoffwechselwege.

3. Enzyme und Stoffwechselwege lassen sich mittels Metadaten gruppieren und typische Auffälligkeiten feststellen. Dies wird mit einigen Beispielen belegt (Kapitel 3.4, Kapitel 3.5 und Kapitel 3.6) und stellt unter anderem diverse konforme Reaktionsschritte des Entner-Doudoroff-Weges und deren enzymatische Katalyse in Zusammenhang mit verschiedenen Lebensräumen der Archaeen und Rhodobacteraceen. Dieser Stoffwechselweg stellt eine wichtige Alternative zur Energiegewinnung durch den Abbau von Zuckern neben der Glycolyse dar und evolvierte in den entsprechenden Organismen unterschiedliche enzymatische Funktionen zur Umsetzung. Ebenso werden typische Verknüpfungen von Reaktionsmechanismen zur Entgiftung toxischer, jedoch im Habitat häufig vorkommender Substrate erkannt. Demgemäß werden Enzymklassen, die zur Quecksilberentgiftung notwendig sind, gehäuft bei acidophilen Archaeen und marinen Rhodobacteraceen, die im Sediment leben, annotiert. Diese zumeist sauren Bedingungen fördern die Löslichkeit des toxischen Quecksilbers und begründen die Notwendigkeit der Organismen, darauf reagieren zu können. Im gleichen Zusammenhang stehen beispielsweise die gezeigten Abbauwege von Halogenverbindungen bei marinen Rhodobacteraceen und besondere Stoffwechselwege acidophiler Archaeen, die ein toxisches Insektizid namens Parathion und dessen Derivate darüber hinaus als Kohlenstoffquelle verwenden. Die Ursache ist womöglich in einer parallelen Evolution der Paraoxinasen mit Enzymen der Klasse der Laktonasen zur Hydrolyse von Lactonen zu finden. Dennoch zeigt dieses Beispiel deutlich die metadatentypische Entwicklung von verschiedenen Enzymklassen der Organismen. Genauso steht ausschließlich marinen Bakterien der Rhodobacteraceen mit der Ectoin-Biosynthese ein Osmoprotektant und Schutzfaktor vor UV-Licht zur Verfügung, der in weiterführenden Untersuchungen interessante Aspekte verheißt. Überdies wirkt sich der symbiontische Lebensstil mit den Algen auf den Metabolismus dahingehend aus, dass etwa Phytohormone wie Indol-3-acetat synthetisiert werden und dazu dienen, die Stoffproduktion der Wirte ihrem Bedarf entsprechend zu induzieren. Ein ebenso eindeutiges Unterscheidungskriterium der Archaeen sind die Stoffwechselwege der Methanogenese. Dementsprechend verfügen Vertreter aus verschiedenen extremen Habitaten über diverse enzymatische Funktionen, um CO₂ zu fixieren und anaerob Biomasse zur Energiegewinnung zu verstoffwechseln. Es werden jedoch auch

vermeintlich untypische Funktionen in diesem Zusammenhang den Rhodobacteraceen zugeordnet. Die Zeiten knapper Nährstoffe und demzufolge oligotropher Bedingungen nach der Algenblüte in den Ozeanen stellt dem Metabolismus dieser Bakterien spezifische Anforderungen zur Biomasseproduktion und Energiegewinnung. Dabei weisen teils noch zu postulierende Stoffwechselwege zur Verwertung von Phosphaten und anderen methylierten Verbindungen auf eine messbare Methanproduktion im sauerstoffreichen Oberflächenwasser hin, während die vorliegende Arbeit marintypische enzymatische Funktionen in diesem Kontext aufzeigt.

Mit Hilfe dieser und weiterer dargestellter Beispiele lassen sich entsprechende Verknüpfungen der Organismen und speziell ihrem Metabolismus zu den Metadaten nachvollziehen und Enzymklassen und Stoffwechselwege eingruppiieren und interpretieren.

4. Die Untersuchung phylogenetischer Kladen auf Enzym-, bzw. Stoffwechselwebebene kann durch die Zusammenstellung von wiederkehrenden Stoffwechselweglücken und gehäuft vorkommenden Enzymen ohne bisherige Zuordnung neue Enzymvorschläge und neue Stoffwechselwegansätze hervorbringen. Die Problematik, dass einerseits oft Enzyme noch in keinem beschriebenen metabolischen Zusammenhang in der untersuchten Domäne stehen und andererseits diese Betrachtungen direkt von der Quantität der herangezogenen Stoffwechselwegdatenbank abhängen, wird in den Kapiteln 3.7 und 3.8 deutlich. Nichtsdestotrotz ermöglicht zum Beispiel die Zusammenstellung von annotierten Enzymen der Rhodobacteraceen die teilweise Abbildung einer Reaktionskette eines bekannten Stoffwechselweges zur Resistenz gegenüber dem Antibiotikum Polymyxin, während bei dieser Bakterienfamilie noch einzelne Reaktionsschritte in der Literatur weiterer Nachweise bedürfen. Ganz ähnlich verhält es sich mit der Evolution des Histidin Abbaus bei den Archaeen. Die annotierten enzymatischen Funktionen weisen bei den Eukaryoten auf zumindest funktionell analoge Stoffwechselwege hin, die bei den Archaeen allerdings weiter untersucht werden müssen.

Sehr häufig fallen in beide untersuchten Organismengruppen annotierte Enzymklassen mit einer Verbindung zu regulatorischen Mechanismen aus Signal- und Transportwegen. Diese sind weniger dazu notwendig, um den Metabolismus eines Organismus nachzuvollziehen, stellen jedoch für die Lebewesen essentielle und unverzichtbare Funktionen zur Verfügung.

5. Die Archaeen brachten im Laufe der Evolution eigens entstandene Enzyme hervor, die sich nicht zwangsweise in der Funktion zu den Bacteria oder auch den Eukaryota unterscheiden, sich jedoch sequenzhomolog deutlich distanzieren (Kapitel 3.4). Dies wird vor allem deutlich bei der Betrachtung des archaeentypischen Aufbaus der Zellmembran und der dazu notwendigen Enzyme zur Katalyse der Synthese besonders stabiler Verbindungen und langkettiger Seitenketten aus Isoprenoideinheiten. Ebenfalls sind besondere DNA-/RNA Schutz- und Reparaturmechanismen domärentypisch und notwendig für die

Archaeen, um bei derart harschen Umweltbedingungen zu überleben, die in ihren Habitaten vorherrschen. Der Kontext und die Bedeutung einiger herausgestellter Enzyme des Verfahrens sind allerdings hier noch nicht eindeutig nachzuvollziehen. Dennoch ist häufig eine breite Substratspezifität domänenspezifischer Enzyme der Archaeen gemeinsam, was ihnen womöglich mit einem weit aufgestellten Metabolismus das Leben in kargen und extremen Habitaten ermöglicht.

Alles in allem bestätigen die Leitthesen, dass die nachvollzogenen Verknüpfungen des Metabolismus von Organismen zu den jeweiligen Metadaten interessante Enzymfunktionen und Stoffwechselwege aggregieren und Zusammenhänge und Aspekte für weiterführende Analysen darbieten. Dementsprechend werden Interpretationsansätze folgen, indem der Pool an Metadaten erweitert wird. Diese Sammlung an Organismusinformationen kann durch zusätzliche Literaturrecherche erweitert und verifiziert werden. Das Ziel sollte ein automatisiertes Textmining von Organismusbeschreibungen aus verschiedensten Quellen sein, um jederzeit erweiterbare Gruppierungen und Kategorisierungen der Lebensweise, der Habitate und der Umweltbedingungen von untersuchten Analysegruppen gegenüberzustellen und spezifische enzymatische Funktionen und Stoffwechselwege zu erkennen. Dabei ist ferner zu untersuchen, ob und welche Differenzen in ausgewählten kleineren Kladen und Gruppierungen zu finden sind. Beispielsweise haben sich interessante Differenzen in Spezies der Ordnung der Sulfolobales aufgezeigt, die es weiter zu untersuchen gilt. Letztendlich sind diese Verfahren auch in einer Anwendung über den gesamten „Tree of Life“ denkbar, um den Metabolismus einerseits auf der Domänen- und der Speziesebene abzugrenzen und andererseits evolvierte Stoffwechselwege verschiedener Organismen zu gruppieren. Dies ist eine große Hilfe für die zukünftige Systembiologie und die Erstellung metabolischer Netzwerke und Modelle, vor allem um Fehlinterpretationen vorzubeugen.

Der Zuordnung von Stoffwechselwegen geht immer die Annotation von Enzymklassen zu den Genomen voraus. Dies ist Ausgangspunkt aller weiterführenden Analysen und von fundamentaler Bedeutung. Infolgedessen wird eine Aufgabe des EnzymeDetectors fortwährend sein, möglichst eindeutige Konfidenzen für enzymatische Annotationen durch die Einbindung weiterer Datenbanken und eigener Algorithmen zu liefern. Die Funktionsvorhersage bei Eukaryota stellt ein ebenso anzustrebendes Ziel dar. Dies wird Quervergleiche vornehmlich von archaealen Modellen erleichtern und interessante Analysen einer weiteren Domäne neben den Archaeen und Bakterien ermöglichen. Ein weiterer wichtiger Punkt ist die Zuordnung von Stoffwechselwegen auf Grundlage der annotierten Enzymklassen. Die in dieser Arbeit diesbezüglich getroffenen definitiven Aussagen durch die Unterscheidung von Schlüsselenzymen und alternativen Reaktionszweigen in Pathways sind durch das erstellte Programm zur Auswertung der Reaktionsketten (*reaction-layout* von MetaCyc) möglich. Dieses Programm soll erweitert werden, sodass selbst komplexeste Zyklen und sogenannte Superpathways abzubilden sind. Allerdings ist wie in Kapitel 3.6 zusammengefasst, ein Layout zu entwerfen, welches beispielsweise strikt

katabolische und anabolische Reaktionswege trennt und so eindeutige Aussagen zur Verstoffwechslung von Substraten ermöglicht. Dazu sollten ebenso weitere Datenbanken, wie zum Beispiel die entstehenden Stoffwechselwegkarten der BRENDA, hinzugezogen werden, wie zur Einordnung regulatorischer Enzyme von Transport- und Signalwegen, die großes Potential zur Differenzierung der Organismen mit Bezug auf Metadaten aufweisen.

Abschließend bleibt zu erwähnen, dass die vorgestellten Methoden und die geschaffene Grundlage durch den EnzymeDetector die Modellerstellung in der Systembiologie vereinfachen, aber auch optimieren und dazu anhalten, bisher grundlegende Annahmen von Funktionszuordnungen zu hinterfragen. In Zukunft können die Metadaten der Organismen hier eine entscheidende Hilfestellung geben. Ferner kann dies im Kontext von Stoffwechselwegen nachvollzogen und so weitere Zusammenhänge postuliert werden. Allerdings kann die Aufspaltung eines komplexen Metabolismus von Organismen in einzelne mehr oder weniger voneinander abhängige Stoffwechselwege immer nur Anhaltspunkt für das Verständnis unsererseits sein, um derartige Modelle zu erstellen. Die Details zur Differenzierung und zum Verständnis von Spezies sind folgerichtig auch immer in der Gesamtheit der metabolischen Funktionen in Form von Enzymklassen, regulierten Abläufen und Mechanismen und Beziehungen zu den Umwelteinflüssen zu suchen.

Danksagung

*„Auch in Wissenschaften kann man eigentlich nichts wissen.
Es will immer getan sein.“*

Johann Wolfgang von Goethe

Diese Arbeit ist mit verschiedenster Unterstützung vieler Menschen zustande gekommen. Auf diesem Wege möchte ich allen danken, die mir im Verlauf dieser Arbeit jederzeit hilfreich zur Seite standen.

Allen voran gilt mein besonderer Dank Herrn Prof. Dr. Schomburg für die äußerst interessante Themenstellung, die Gelegenheit das Feld der Systembiologie und der Bioinformatik tiefgreifend zu erforschen und für all die hilfreichen Diskussions- und Erörterungsgespräche zu jeder Zeit. Des Weiteren möchte ich Prof. Dr. Jahn und Prof. em. Dr. Ehrich für die Prüfung und Bewertung dieser Arbeit herzlich danken.

Wie Johann Wolfgang von Goethe schon sagte, wird unsere Wissenschaft uns fortwährend Fragen stellen, die wir nicht allein beantworten können und praktisch nachvollziehen müssen. Umso wichtiger sind die Zusammenarbeit und die fachliche Unterstützung, sowie ein motivierendes Arbeitsumfeld, um Tag für Tag diesen Fragen nachzugehen. An nichts mangelte es während meiner Arbeit in der Arbeitsgruppe des Instituts für Bioinformatik und Biochemie an der Technischen Universität Carolo-Wilhelmina zu Braunschweig.

In diesem Sinne danke ich außerordentlich Frau Dr. Antje Chang und Frau Dr. Ida Schomburg, die mir zu jederzeit mit fachlicher Kompetenz und Rat zur Seite standen, die es verstehen zu motivieren und darüber hinaus unermüdlich für ein angenehmes Arbeitsklima mit freundlichster Unterstützung sorgen.

Dem gesamten BRENDA-Team möchte ich für die Unterstützung und freundliche Zusammenarbeit danken. Der gesamten Arbeitsgruppe gilt mein Dank für die hilfreichen Diskussionen und das nette Arbeitsklima. Vornehmlich Frau Dipl.-Biomath. Lisa Jeske, Herrn Msc Biologie Helge Stark und den Modellierern der Arbeitsgruppe möchte ich für die enge Zusammenarbeit, das so wichtige Feedback und all die glücklichen Momente der Freundschaft danken. Für die experimentelle Unterstützung und biochemischen Erläuterungen gilt mein Dank insbesondere Frau Dr. Kerstin Schmidt-Hohagen, Frau Dipl.-Lebensmittelchemikerin Sarah Kleist und dem Laborteam des Instituts.

Daneben möchte in ganz besonderer Weise meiner Familie danken, die mir diesen Weg ebnete, meine Interessen in den jüngsten Jahren weckte und ein liebevolles Umfeld, sowie einen unterstützenden Hintergrund zu jeder Zeit schafft. Zu guter Letzt gilt mein herzlichster Dank Frau Dipl. Bioinf. Sandra Placzek für die technische und fachliche Unterstützung, all die hilfreichen Erörterungen und Diskussionen logischer Fragestellungen und die entscheidenden Hinweise auf dem Weg dieser Arbeit zu jeder Tageszeit und an jedem Ort. Besonders danken möchte ich ihr weiterhin für ihre unendliche Geduld, die sie mir ebenso entgegenbrachte, wie sie es versteht, unaufhörlich zu motivieren, zu inspirieren und dabei hilft zu realisieren. Diese so wichtigen Aspekte der Wissenschaft leben vom Glück und Wohlbefinden der Wissenschaftler, welches ich mit ihr nicht besser hätte finden können.

Symbole und Abkürzungsverzeichnis

α	griechisch: alpha
β	griechisch: beta
16S rRNA	Ribonukleinsäure mit einem Sedimentationskoeffizienten von 16 der kleinen Ribosomenuntereinheit
ABC Transporter	<i>ATP-binding cassette</i> -Transporter
ADP	Adenosindiphosphat
AMENDA	<i>Automatic Mining of ENzyme DATA</i>
ATP	Adenosintriphosphat
BLAST	<i>Basic Local Alignment Search Tool</i>
BRENDA	BRAunschweiger <i>ENzyme DAtabase</i>
BrEPS	Braunschweiger <i>Enzyme Pattern Search</i>
CO₂	Kohlenstoffdioxid
CoA	Coenzym A
CSV	<i>Comma seperated values</i>
DB	<i>DataBase</i>
DNA	Desoxyribonukleinsäure
EC	<i>Enzyme Commission</i>
E-Value	<i>Expected-Value</i>
FAD	<i>Flavin Adenine Dinucleotide</i> (oxidiert)
FADH	<i>Flavin Adenine Dinucleotide</i> (reduziert)
FTP	<i>File Transfer Protocol</i>
GI	<i>Gene Identifier</i>
HMM	<i>Hidden Markov Model</i>
HTML	<i>Hypertext Markup Language</i>
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
Klade	monophyletische Gruppe
NAD⁺	<i>Nicotinamide Adenine Dinucleotide</i> (oxidiert)
NADH	<i>Nicotinamide Adenine Dinucleotide</i> (reduziert)
NADP⁺	<i>Nicotinamide Adenine Dinucleotide Phosphate</i> (oxidiert)
NADPH	<i>Nicotinamide Adenine Dinucleotide Phosphate</i> (reduziert)
NCBI	<i>National Center for Biotechnology Information</i>
RefSeq	<i>Reference Sequence</i>
RNA	Ribonukleinsäure
SQL	<i>Structured Query Language</i>
TCA	Tricarbonsäurezyklus/Citratzyklus
TrEMBL	<i>Translated EMBL Nucleotide Sequence Data Library</i>
UniParc	<i>UniProt Archive</i>
UniProt	<i>Universal Protein Resource</i>
UniProtKB	<i>UniProt Knowledgebase</i>
UniRef	<i>UniProt Reference Clusters</i>

Abbildungsverzeichnis

Abbildung 1:	Schema Verfahren I (Archaeen)	35
Abbildung 2:	Schema Verfahren II (Analysegruppen)	39
Abbildung 3:	Vergleich des Zeitmanagements	41
Abbildung 4:	EnzymeDetector Organismenglossar	46
Abbildung 5:	Neues Webinterface des EnzymeDetector	47
Abbildung 6:	TCA Cycle I	49
Abbildung 7:	Archaeentypisches Enzym	71
Abbildung 8:	Bakterientypisches Enzym	72
Abbildung 9:	Neighbor Joining-Baum der 16S rRNA Sequenzen der Rhodobacteraceen	102
Abbildung 10:	Neighbor Joining-Baum der binären Enzym-Vektoren der Rhodobacteraceen	104
Abbildung 11:	Neighbor Joining-Baum der 16S rRNA Sequenzen der Archaeen	106
Abbildung 12:	Neighbor Joining-Baum der binären Enzym-Vektoren der Archaeen	108
Abbildung 13:	Neighbor Joining-Baum der binären Pathway-Vektoren der Rhodobacteraceen	110
Abbildung 14:	Neighbor Joining-Baum der binären Pathway-Vektoren der Archaeen	113

Tabellenverzeichnis

Tabelle 1:	Scoring-Schema des EnzymeDetector im Original	9
Tabelle 2:	Typisierte Metadaten und ihre Kategorien in den Projektorganismengruppen	28
Tabelle 3:	Spezifikationen der HPC-Cluster Topologie	30
Tabelle 4:	Scoring-Schema des EnzymeDetector	44
Tabelle 5:	Das BRENDA-Scoring-Schema	45
Tabelle 6:	Enzym-Evolution der Archaeen: pH-Wert	51
Tabelle 7:	Enzym-Evolution der Archaeen: Temperatur	58
Tabelle 8:	Enzym-Evolution der Archaeen: Sauerstoffmetabolismus	63
Tabelle 9:	Enzym-Evolution der Archaeen: Habitat	65
Tabelle 10:	Enzym-Evolution der Archaeen: Halotoleranz	67
Tabelle 11:	Enzym-Evolution der Archaeen: methanogene Archaeen	69
Tabelle 12:	Domänentypische Enzyme	73
Tabelle 13:	Metadatentypische Enzyme der Archaeen	76
Tabelle 14:	Metadatentypische Enzyme der Rhodobacteraceen	80
Tabelle 15:	Metadatentypische Stoffwechselwege der Archaeen	84
Tabelle 16:	Metadatentypische Stoffwechselwege der Rhodobacteraceen	87
Tabelle 17:	Gruppentypische Pathway-Lücken	91
Tabelle 18:	Gruppentypische Enzyme ohne Pathwayzuordnung	94
Tabelle 19:	Core-/Pan-„Enzymom“ der Projektgruppen	99

Anhang

Anhang 1: Datentabellen des EnzymeDetector-Ansatzes der Rhodobacteraceen.

Tabellenblätter:

1. *rho_ec*: Metadaten- und Enzymmatrix der Rhodobacteraceen.
2. *rho_pw*: Metadaten- und Stoffwechselwegematrix der Rhodobacteraceen.
3. *rho_gaps*: Daten der gezählten Lücken durch häufig nicht annotierte Enzyme.
4. *rho_new_pws*: Häufigkeit von Enzymen ohne Pathwayzuordnung.
5. *core_pan_ec*: Auswertung des *Core*- und *Pan*-„Enzynoms“.
6. *rho_ec_raw* und *rho_pw_raw*: Rohdaten vom Python-Skript CSV-Output.
7. *ecs-pws* und *pws-ecs*: Zuordnungen von Enzymen zu Pathways.

Siehe CD-ROM: rho.xlsx

Anhang 2: Datentabellen des EnzymeDetector-Ansatzes der Archaeen.

Tabellenblätter:

1. *arch_ec*: Metadaten- und Enzymmatrix der Archaeen.
2. *arch_pw*: Metadaten- und Stoffwechselwegematrix der Archaeen.
3. *arch_gaps*: Daten der gezählten Lücken durch häufig nicht annotierte Enzyme.
4. *arch_new_pws*: Häufigkeit von Enzymen ohne Pathwayzuordnung.
5. *core_pan_ec*: Auswertung des *Core*- und *Pan*-„Enzynoms“.
6. *arch_ec_raw* und *arch_pw_raw*: Rohdaten vom Python-Skript CSV-Output.
7. *ecs-pws* und *pws-ecs*: Zuordnungen von Enzymen zu Pathways.

Siehe CD-ROM: arch.xlsx

Anhang 3: Datentabellen des BLAST-Ansatzes der Archaeen.

Datenmatrix der Häufigkeit der Enzyme in den Metadaten-Kategorien.

Siehe CD-ROM: Analysis_MetaData.xlsx

Anhang 4: Datentabellen des BLAST-Ansatzes der Archaeen zur Auswertung der E-Value-Statistik.

Berechnungsmatrix der Statistik und Filtern der domänentypischen Enzyme.

Siehe CD-ROM: Analysis_eValue.xlsx

Anhang 5: Quellcode des EnzymeDetector, des Pathway-Layout, der Skripte und die Datenbanken

Siehe CD-ROM: Ordner Sourcecode

Literaturverzeichnis

1. Yoshida S, Hiraga K, Takehana T, Taniguchi I, Yamaji H, Maeda Y, Toyohara K, Miyamoto K, Kimura Y, Oda K: **A bacterium that degrades and assimilates poly(ethylene terephthalate)**. *Science* 2016;**351**(6278):1196–9, doi:10.1126/science.aad6359.
2. Berg JM, Tymoczko JL, Gatto GJ, Stryer L. *Biochemistry*. 8th ed.: W.H. Freeman/Macmillan; 2015.
3. Strimbu K, Tavel JA: **What are biomarkers?** *Current Opinion in HIV and AIDS* 2010;**5**(6):463–6, doi:10.1097/COH.0b013e32833ed177.
4. Mavromoustakos T, Durdagi S, Koukoulitsa C, Simcic M, Papadopoulos MG, Hodoscek M, Grdadolnik SG: **Strategies in the rational drug design**. *Current medicinal chemistry* 2011;**18**(17):2517–30.
5. Nielsen J: **Metabolic engineering**. *Applied microbiology and biotechnology* 2001;**55**(3):263–83.
6. Chuang H, Hofree M, Ideker T: **A Decade of Systems Biology**. *Annu. Rev. Cell Dev. Biol.* 2010;**26**(1):721–44, doi:10.1146/annurev-cellbio-100109-104122.
7. Kitano H: **Systems biology: a brief overview**. *Science (New York, N.Y.)* 2002;**295**(5560):1662–4, doi:10.1126/science.1069492.
8. Kohlstedt M, Becker J, Wittmann C: **Metabolic fluxes and beyond—systems biology understanding and engineering of microbial metabolism**. *Applied microbiology and biotechnology*. - Berlin Springer. - Heidelberg [u.a.] 2010;**88**(5):1065–75.
9. Price ND, Reed JL, Palsson BØ: **Genome-scale models of microbial cells: evaluating the consequences of constraints**. *Nature reviews. Microbiology* 2004;**2**(11):886–97, doi:10.1038/nrmicro1023.
10. Price ND: **Genome-Scale Models for Microbial Factories**. *Industrial biotechnology*. - New Rochelle 2013;**9**(4):177–8.
11. Ulas T, Riemer SA, Zaparty M, Siebers B, Schomburg D, Parkinson J: **Genome-Scale Reconstruction and Analysis of the Metabolic Network in the Hyperthermophilic Archaeon Sulfolobus Solfataricus**. *PloS one* 2012;**7**(8):e43401, doi:10.1371/journal.pone.0043401.
12. Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110**. *Applied and Environmental Microbiology* 1994;**60**(10):3724–31.
13. Horton HR, Moran LA, Scrimgeour KG, Perry MD, Rawn JD, Biele C. *Biochemie*. 4th ed.: Pearson Studium; 2008.
14. Campbell NA, Reece JB, Held A, Markl J (Eds). *Biologie*. 6th ed.: Pearson Studium; 2008.
15. Lee D, Chung BKS, Yusufi FN, Selvarasu S: **In silico genome-scale modeling and analysis for identifying anti-tubercular drug targets**. *Drug Dev. Res.* 2011;**72**(2):121–9, doi:10.1002/ddr.20408.
16. Becker J, Zelder O, Häfner S, Schröder H, Wittmann C: **From zero to hero—Design-based systems metabolic engineering of Corynebacterium**

- glutamicum for l-lysine production.** *Metabolic Engineering* 2011;**13**(2):159–68, doi:10.1016/j.ymben.2011.01.003.
17. Zhang Y, Thiele I, Weekes D, Li Z, Jaroszewski L, Ginalski K, Deacon AM, Wooley J, Lesley SA, Wilson IA *et al.*: **Three-Dimensional Structural View of the Central Metabolic Network of *Thermotoga maritima*.** *Science* 2009;**325**(5947):1544–9, doi:10.1126/science.1174671.
 18. Shahzad K, J. Loor J: **Application of Top-Down and Bottom-up Systems Approaches in Ruminant Physiology and Metabolism.** *CG* 2012;**13**(5):379–94, doi:10.2174/138920212801619269.
 19. Çakır T, Khatibipour MJ: **Metabolic Network Discovery by Top-Down and Bottom-Up Approaches and Paths for Reconciliation.** *Front. Bioeng. Biotechnol.* 2014;**2**(Suppl. 1):e1002518, doi:10.3389/fbioe.2014.00062.
 20. Chang A, Schomburg I, Placzek S, Jeske L, Ulbrich M, Xiao M, Sensen CW, Schomburg D: **BRENDA in 2015. Exciting developments in its 25th year of existence.** *Nucleic Acids Research* 2015;**43**(D1):D439–D446, doi:10.1093/nar/gku1068.
 21. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A *et al.*: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases.** *Nucl. Acids Res.* 2013;**42**(D1):D459–D471, doi:10.1093/nar/gkt1103.
 22. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M: **KEGG as a reference resource for gene and protein annotation.** *Nucleic Acids Res* 2016;**44**(D1):D457–D462, doi:10.1093/nar/gkv1070.
 23. Quester S, Schomburg D: **EnzymeDetector. An integrated enzyme function prediction tool and database.** *BMC Bioinformatics* 2011;**12**(1):376, doi:10.1186/1471-2105-12-376.
 24. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic acids research: NAR.* - Oxford Oxford Univ. Press. - 2000;**28**(1):27–30.
 25. Tatusova T, Ciufu S, Fedorov B, O'Neill K, Tolstoy I: **RefSeq microbial genomes database. New representation and annotation strategy.** *Nucleic Acids Research* 2015;**43**(7):3872, doi:10.1093/nar/gkv278.
 26. Bethesda MD. *The NCBI handbook.* Available at: <http://www.ncbi.nlm.nih.gov/books/NBK21091/>; 2002.
 27. Frishman D, Mokrejs M, Kosykh D, Kastenmüller G, Kolesov G, Zubrzycki I, Gruber C, Geier B, Kaps A, Albermann K *et al.*: **The PEDANT genome database.** *Nucleic Acids Research* 2003;**31**(1):207–11.
 28. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BO: **A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism--2011.** *Molecular Systems Biology* 2011;**7**(1):535, doi:10.1038/msb.2011.65.
 29. Riemer A. *metano: An open-source toolbox for the reconstruction and analysis of genome-scale metabolic models*; 2013.
 30. INTERNATIONAL UNION OF BIOCHEMISTRY AND MOLECULAR BIOLOGY. *IUBMB Nomenclature Committee.* Available at: <http://www.chem.qmul.ac.uk/iubmb/>; 2016.

- 31.Jiang W, Bond JS: **Families of metalloendopeptidases and their relationships.** *FEBS Letters* 1992;**312**(2-3):110–4, doi:10.1016/0014-5793(92)80916-5.
- 32.**UniProt. A hub for protein information.** *Nucleic Acids Research* 2015;**43**(D1):D204-D212, doi:10.1093/nar/gku989.
- 33.Seyffert W, Balling R (Eds). *Lehrbuch der Genetik*. 2nd ed.: Spektrum Akad. Verl; 2003.
- 34.James CL, Viola RE: **Production and Characterization of Bifunctional Enzymes. Domain Swapping To Produce New Bifunctional Enzymes in the Aspartate Pathway †.** *Biochemistry* 2002;**41**(11):3720–5, doi:10.1021/bi015909o.
- 35.Moore Bd: **Bifunctional and moonlighting enzymes. Lighting the way to regulatory control.** *Trends in Plant Science* 2004;**9**(5):221–8, doi:10.1016/j.tplants.2004.03.005.
- 36.Bannert C, Welfle A, dem Spring C aus, Schomburg D: **BrEPS. A flexible and automatic protocol to compute enzyme-specific sequence profiles for functional annotation.** *BMC Bioinformatics* 2010;**11**(1):589, doi:10.1186/1471-2105-11-589.
- 37.Hartman AL, Norais C, Badger JH, Delmas S, Haldenby S, Madupu R, Robinson J, Khouri H, Ren Q, Lowe TM *et al.*: **The Complete Genome Sequence of Haloferax volcanii DS2, a Model Archaeon.** *PloS one* 2010;**5**(3):e9605, doi:10.1371/journal.pone.0009605.
- 38.Garrett RA, Klenk H (Eds). *Archaea: Evolution, physiology, and molecular biology*. Blackwell Publ; 2007.
- 39.Case RJ, Boucher Y, Dahllöf I, Holmstrom C, Doolittle WF, Kjelleberg S: **Use of 16S rRNA and rpoB Genes as Molecular Markers for Microbial Ecology Studies.** *Applied and Environmental Microbiology* 2006;**73**(1):278–88, doi:10.1128/AEM.01177-06.
- 40.Zhang J, Zhu B, Xu C, Ding X, Li J, Zhang X, Lu Z: **[Strategy of selecting 16S rRNA hypervariable regions for metagenome-phylogenetic marker genes based analysis].** *Ying yong sheng tai xue bao = The journal of applied ecology / Zhongguo sheng tai xue xue hui, Zhongguo ke xue yuan Shenyang ying yong sheng tai yan jiu suo zhu ban* 2015;**26**(11):3545–53.
- 41.Griffiths AJF. *Introduction to genetic analysis*. 8th ed.: Freeman; 2005.
- 42.Kloda A, Martinac B: **Mechanosensitive channel of Thermoplasma, the cell wall-less archaea: cloning and molecular characterization.** *Cell biochemistry and biophysics* 2001;**34**(3):321–47, doi:10.1385/CBB:34:3:321.
- 43.Pearson A, Pi Y, Zhao W, Li W, Li Y, Inskeep W, Perevalova A, Romanek C, Li S, Zhang CL: **Factors Controlling the Distribution of Archaeal Tetraethers in Terrestrial Hot Springs.** *Applied and Environmental Microbiology* 2008;**74**(11):3523–32, doi:10.1128/AEM.02450-07.
- 44.Boyd ES, Hamilton TL, Wang J, He L, Zhang CL: **The Role of Tetraether Lipid Composition in the Adaptation of Thermophilic Archaea to Acidity.** *Front. Microbiol.* 2013;**4**, doi:10.3389/fmicb.2013.00062.
- 45.Herzog B, Wirth R: **Swimming Behavior of Selected Species of Archaea.** *Applied and Environmental Microbiology* 2012;**78**(6):1670–4, doi:10.1128/AEM.06723-11.
- 46.URICH T, BANDEIRAS TM, LEAL SS, RACHEL R, ALBRECHT T, ZIMMERMANN P, SCHOLZ C, TEIXEIRA M, GOMES CM, KLETZIN A: **The**

- sulphur oxygenase reductase from *Acidianus ambivalens* is a multimeric protein containing a low-potential mononuclear non-haem iron centre. *Biochem. J.* 2004;**381**(1):137–46, doi:10.1042/BJ20040003.
47. Soppa J: **Evolutionary advantages of polyploidy in halophilic archaea.** *Biochem. Soc. Trans.* 2013;**41**(1):339–43, doi:10.1042/BST20120315.
 48. Gudhka RK, Neilan BA, Burns BP: **Adaptation, Ecology, and Evolution of the Halophilic Stromatolite Archaeon *Halococcus hamelinensis* Inferred through Genome Analyses.** *Archaea* 2015;**2015**(1):1–11, doi:10.1155/2015/241608.
 49. Falb M, Müller K, Königsmaier L, Oberwinkler T, Horn P, Gronau S von, Gonzalez O, Pfeiffer F, Bornberg-Bauer E, Oesterhelt D: **Metabolism of halophilic archaea.** *Extremophiles* 2008;**12**(2):177–96, doi:10.1007/s00792-008-0138-x.
 50. Bolhuis H, Palm P, Wende A, Falb M, Rampp M, Rodriguez-Valera F, Pfeiffer F, Oesterhelt D: **The genome of the square archaeon *Haloquadratum walsbyi* life at the limits of water activity.** *BMC Genomics* 2006;**7**(1):169, doi:10.1186/1471-2164-7-169.
 51. Tajima K, Nagamine T, Matsui H, Nakamura M, Aminov RI: **Phylogenetic analysis of archaeal 16S rRNA libraries from the rumen suggests the existence of a novel group of archaea not associated with known methanogens.** *FEMS microbiology letters* 2001;**200**(1):67–72.
 52. Zuo G, Xu Z, Hao B: **Phylogeny and Taxonomy of Archaea. A Comparison of the Whole-Genome-Based CVTree Approach with 16S rRNA Sequence Analysis.** *Life* 2015;**5**(1):949–68, doi:10.3390/life5010949.
 53. Huber H, Stetter KO: **Thermoplasmatales.** In: *The Prokaryotes. The Prokaryotes. Edited by Dworkin, M, Falkow, S, Rosenberg, E, Schleifer, K, Stackebrandt, E: Springer New York; 2006:101–12.*
 54. Yu D, Kurola JM, Lähde K, Kymäläinen M, Sinkkonen A, Romantschuk M: **Biogas production and methanogenic archaeal community in mesophilic and thermophilic anaerobic co-digestion processes.** *Journal of Environmental Management* 2014;**143**:54–60, doi:10.1016/j.jenvman.2014.04.025.
 55. Ilk N, Egelseer EM, Sleytr UB: **S-layer fusion proteins—construction principles and applications.** *Current Opinion in Biotechnology* 2011;**22**(6):824–31, doi:10.1016/j.copbio.2011.05.510.
 56. Sleytr UB, Sára M: **Bacterial and archaeal S-layer proteins: structure-function relationships and their biotechnological applications.** *Trends in biotechnology* 1997;**15**(1):20–6, doi:10.1016/S0167-7799(96)10063-9.
 57. Sleytr UB, Schuster B, Egelseer E, Pum D: **S-layers. Principles and applications.** *FEMS Microbiol Rev* 2014;**38**(5):823–64, doi:10.1111/1574-6976.12063.
 58. Ishino S, Ishino Y: **DNA polymerases as useful reagents for biotechnology – the history of developmental research in the field.** *Front. Microbiol.* 2014;**5**(e51):2216, doi:10.3389/fmicb.2014.00465.
 59. Kong H, Kucera RB, Jack WE: **Characterization of a DNA polymerase from the hyperthermophile archaea *Thermococcus litoralis*. Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities.** *The Journal of biological chemistry* 1993;**268**(3):1965–75.

60. Imanaka T, Atomi H: **Catalyzing "hot" reactions: enzymes from hyperthermophilic Archaea.** *Chem. Record* 2002;**2**(3):149–63, doi:10.1002/tcr.10023.
61. Wagner-Döbler I, Biebl H: **Environmental Biology of the Marine Roseobacter Lineage.** *Annu. Rev. Microbiol.* 2006;**60**(1):255–80, doi:10.1146/annurev.micro.60.080805.142115.
62. Moran MA, Belas R, Schell MA, Gonzalez JM, Sun F, Sun S, Binder BJ, Edmonds J, Ye W, Orcutt B *et al.*: **Ecological Genomics of Marine Roseobacters.** *Applied and Environmental Microbiology* 2007;**73**(14):4559–69, doi:10.1128/AEM.02580-06.
63. Brinkhoff T, Giebel H, Simon M: **Diversity, ecology, and genomics of the Roseobacter clade. A short overview.** *Arch Microbiol* 2008;**189**(6):531–9, doi:10.1007/s00203-008-0353-y.
64. Elifantz H, Horn G, Ayon M, Cohen Y, Minz D: **Rhodobacteraceae are the key members of the microbial community of the initial biofilm formed in Eastern Mediterranean coastal seawater.** *FEMS Microbiol Ecol* 2013;**85**(2):348–57, doi:10.1111/1574-6941.12122.
65. Krohn-Molt I, Wemheuer B, Alawi M, Poehlein A, Gullert S, Schmeisser C, Pommerening-Roser A, Grundhoff A, Daniel R, Hanelt D *et al.*: **Metagenome Survey of a Multispecies and Alga-Associated Biofilm Revealed Key Elements of Bacterial-Algal Interactions in Photobioreactors.** *Applied and Environmental Microbiology* 2013;**79**(20):6196–206, doi:10.1128/AEM.01641-13.
66. Gilbert P, Collier PJ, Brown MR: **Influence of growth rate on susceptibility to antimicrobial agents: biofilms, cell cycle, dormancy, and stringent response.** *Antimicrobial agents and chemotherapy* 1990;**34**(10):1865–8.
67. Buchan A, Gonzalez JM, Moran MA: **Overview of the Marine Roseobacter Lineage.** *Applied and Environmental Microbiology* 2005;**71**(10):5665–77, doi:10.1128/AEM.71.10.5665-5677.2005.
68. Bruhn JB, Gram L, Belas R: **Production of Antibacterial Compounds and Biofilm Formation by Roseobacter Species Are Influenced by Culture Conditions.** *Applied and Environmental Microbiology* 2007;**73**(2):442–50, doi:10.1128/AEM.02238-06.
69. Wagner-Döbler I, Ballhausen B, Berger M, Brinkhoff T, Buchholz I, Bunk B, Cypionka H, Daniel R, Drepper T, Gerdt G *et al.*: **The complete genome sequence of the algal symbiont Dinoroseobacter shibae. A hitchhiker's guide to life in the sea.** *ISME J* 2009;**4**(1):61–77, doi:10.1038/ismej.2009.94.
70. Shieh WY, Lin Y, Jean WD: **Pseudovibrio denitrificans gen. nov., sp. nov., a marine, facultatively anaerobic, fermentative bacterium capable of denitrification.** *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* 2004;**54**(Pt 6):2307–12, doi:10.1099/ijs.0.63107-0.
71. Shearer N, Hinsley AP, Spanning RJ, Spiro S: **PHYSIOLOGY AND METABOLISM - Anaerobic Growth of Paracoccus denitrificans Requires Cobalamin: Characterization of cobK and cobJ Genes.** *Journal of bacteriology: JB / American Society for Microbiology. - Washington* 1999;**181**(22):6907–13.
72. Swingley WD, Sadekar S, Mastrian SD, Matthies HJ, Hao J, Ramos H, Acharya CR, Conrad AL, Taylor HL, Dejesa LC *et al.*: **The Complete Genome**

- Sequence of *Roseobacter denitrificans* Reveals a Mixotrophic Rather than Photosynthetic Metabolism.** *Journal of bacteriology* 2007;**189**(3):683–90, doi:10.1128/JB.01390-06.
73. Rex R, Bill N, Schmidt-Hohagen K, Schomburg D, Maranas CD: **Swimming in Light. A Large-Scale Computational Analysis of the Metabolism of *Dinoroseobacter shibae*.** *PLoS Comput Biol* 2013;**9**(10):e1003224, doi:10.1371/journal.pcbi.1003224.
74. Luo H, Swan BK, Stepanauskas R, Hughes AL, Moran MA: **Evolutionary analysis of a streamlined lineage of surface ocean *Roseobacters*.** *ISME J* 2014;**8**(7):1428–39, doi:10.1038/ismej.2013.248.
75. Allgaier M, Uphoff H, Felske A, Wagner-Dobler I: **Aerobic Anoxygenic Photosynthesis in *Roseobacter* Clade Bacteria from Diverse Marine Habitats.** *Applied and Environmental Microbiology* 2003;**69**(9):5051–9, doi:10.1128/AEM.69.9.5051-5059.2003.
76. Porsby CH, Nielsen KF, Gram L: ***Phaeobacter* and *Ruegeria* Species of the *Roseobacter* Clade Colonize Separate Niches in a Danish Turbot (*Scophthalmus maximus*)-Rearing Farm and Antagonize *Vibrio anguillarum* under Different Growth Conditions.** *Applied and Environmental Microbiology* 2008;**74**(23):7356–64, doi:10.1128/AEM.01738-08.
77. Pujalte MJ, Lucena T, Ruvira MA, Arahal DR, Macián MC: **The Family *Rhodobacteraceae*.** In: *The Prokaryotes. The Prokaryotes: Alphaproteobacteria and Betaproteobacteria.* Edited by Rosenberg, E, DeLong, EF, Lory, S, Stackebrandt, E, Thompson, F: Imprint Springer; 2014:439–512.
78. Luo H, Moran MA: **Evolutionary Ecology of the Marine *Roseobacter* Clade.** *Microbiol. Mol. Biol. Rev.* 2014;**78**(4):573–87, doi:10.1128/MMBR.00020-14.
79. Drüppel K, Hensler M, Trautwein K, Koßmehl S, Wöhlbrand L, Schmidt-Hohagen K, Ulbrich M, Bergen N, Meier-Kolthoff JP, Göker M *et al.*: **Pathways and substrate-specific regulation of amino acid degradation in *Phaeobacter inhibens* DSM 17395 (archetype of the marine *Roseobacter* clade).** *Environ Microbiol* 2014;**16**(1):218–38, doi:10.1111/1462-2920.12276.
80. Wiegmann K, Hensler M, Wöhlbrand L, Ulbrich M, Schomburg D, Rabus R: **Carbohydrate Catabolism in *Phaeobacter inhibens* DSM 17395, a Member of the Marine *Roseobacter* Clade.** *Applied and Environmental Microbiology* 2014;**80**(15):4725–37, doi:10.1128/AEM.00719-14.
81. Kleist S, Ulbrich M, Bill N, Schmidt-Hohagen K, Geffers R, Schomburg D: **Dealing with salinity extremes and nitrogen limitation - an unexpected strategy of the marine bacterium *Dinoroseobacter shibae*.** *Environmental Microbiology* 2016:n/a-n/a, doi:10.1111/1462-2920.13266.
82. Codd EF. *The relational model for database management: Version 2:* Addison-Wesley; 1990.
83. Benson DA: **GenBank.** *Nucleic Acids Research* 2004;**33**(Database issue):D34-D38, doi:10.1093/nar/gki063.
84. Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S *et al.*: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Research* 2009;**37**(Database):D5-D15, doi:10.1093/nar/gkn741.

85. Nakaya A, Katayama T, Itoh M, Hiranuka K, Kawashima S, Moriya Y, Okuda S, Tanaka M, Tokimatsu T, Yamanishi Y *et al.*: **KEGG OC. A large-scale automatic construction of taxonomy-based ortholog clusters.** *Nucleic Acids Research* 2012;**41**(D1):D353-D357, doi:10.1093/nar/gks1239.
86. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al.*: **The Pfam protein families database. Towards a more sustainable future.** *Nucleic Acids Res* 2016;**44**(D1):D279-D285, doi:10.1093/nar/gkv1344.
87. Petrokovski S: **The Blocks database--a system for protein classification.** *Nucleic Acids Research* 1996;**24**(1):197–200, doi:10.1093/nar/24.1.197.
88. Hulo N, Bairoch A, Bulliard V, Cerutti L, Castro E de, Langendijk-Genevaux PS, Pagni M, Sigrist CJA: **The PROSITE database.** *Nucleic Acids Research* 2006;**34**(Database issue):D227-30, doi:10.1093/nar/gkj063.
89. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R *et al.*: **PATRIC, the bacterial bioinformatics database and analysis resource.** *Nucl. Acids Res.* 2013;**42**(D1):D581-D591, doi:10.1093/nar/gkt1099.
90. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M *et al.*: **The RAST Server. Rapid Annotations using Subsystems Technology.** *BMC Genomics* 2008;**9**(1):75, doi:10.1186/1471-2164-9-75.
91. Gremse M, Chang A, Schomburg I, Grote A, Scheer M, Ebeling C, Schomburg D: **The BRENDA Tissue Ontology (BTO). The first all-integrating ontology of all organisms for enzyme sources.** *Nucleic Acids Research* 2010;**39**(Database):D507-D513, doi:10.1093/nar/gkq968.
92. Chang A, Scheer M, Grote A, Schomburg I, Schomburg D: **BRENDA, AMENDA and FRENDA the enzyme information system. New content and tools in 2009.** *Nucleic Acids Research* 2009;**37**(Database):D588-D592, doi:10.1093/nar/gkn820.
93. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of molecular biology* 1990;**215**(3):403–10, doi:10.1016/S0022-2836(05)80360-2.
94. Johnson LS, Eddy SR, Portugaly E: **Hidden Markov model speed heuristic and iterative HMM search procedure.** *BMC Bioinformatics* 2010;**11**(1):431, doi:10.1186/1471-2105-11-431.
95. Selzer PM, Marhöfer RJ, Rohwer A. *Angewandte Bioinformatik: Eine Einführung*: Springer Berlin Heidelberg; 2004.
96. Neron B, Menager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P, Letondal C: **Mobyle. A new full web bioinformatics framework.** *Bioinformatics* 2009;**25**(22):3005–11, doi:10.1093/bioinformatics/btp493.
97. Barrett T, Clark K, Gevorgyan R, Gorelenkov V, Gribov E, Karsch-Mizrachi I, Kimelman M, Pruitt KD, Resenchuk S, Tatusova T *et al.*: **BioProject and BioSample databases at NCBI. Facilitating capture and organization of metadata.** *Nucleic Acids Research* 2011;**40**(D1):D57-D63, doi:10.1093/nar/gkr1163.
98. Tatusova T, Ciufu S, Fedorov B, O'Neill K, Tolstoy I: **RefSeq microbial genomes database. New representation and annotation strategy.** *Nucl. Acids Res.* 2013;**42**(D1):D553-D559, doi:10.1093/nar/gkt1274.

99. Edgar RC: **MUSCLE. Multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Research* 2004;**32**(5):1792–7, doi:10.1093/nar/gkh340.
100. Felsenstein J: **PHYLP – Phylogeny Inference Package (Version 3.2).** *Cladistics* 1989(5):164–6.
101. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Molecular biology and evolution: MBE.* - Oxford University Press. - 2013;**4**(4):406.
102. Wegner P: **A technique for counting ones in a binary computer.** *Commun. ACM* 1960;**3**(5):322, doi:10.1145/367236.367286.
103. Jordan GE, Piel WH: **PhyloWidget. Web-based visualizations for the tree of life.** *Bioinformatics* 2008;**24**(14):1641–2, doi:10.1093/bioinformatics/btn235.
104. Stark H. *PFAM-EC_Families: Zuordnung von Enzymklassen zu PFAM Proteinfamilien.*: Institut für Bioinformatik und Biochemie. Technische Universität Braunschweig; 2015.
105. Jeske L. *MetaCyc Flatfile Parser: Überführung der MetyCyc Datendateien in eine MySQL-Datenbank.*: Institut für Bioinformatik und Biochemie. Technische Universität Braunschweig; 2015.
106. Munaretto C. *Literaturrecherche zu den Metadaten der Archaea.* Institut für Bioinformatik und Biochemie. Technische Universität Braunschweig; 2015.
107. Software in the Public Interest, Inc. *Debian GNU/Linux 7.9 (wheezy)*: GNU General Public License; 2015.
108. Lange T. *FAI v5.0: Fully Automatic Installation.*: GNU General Public License; 2015.
109. Community Projekt. *Son of Grid Engine v8.1.8.*: University of Liverpool; 2015.
110. *WordPress 4.3.* Available at: <https://de.wordpress.org>; 2015.
111. *jQuery.* Available at: <https://jquery.com/>; 2015.
112. *PHP 5.* Available at: <https://secure.php.net/>; 2015.
113. *The World Wide Web Consortium (W3C)*: <https://www.w3.org>; 2016.
114. Oracle Corporation. *MySQL 5.6 Reference Manual.* Available at: <http://dev.mysql.com/doc/refman/5.6/en/>; 2015.
115. Widenius M, Axmark D. *MySQL reference manual: Documentation from the source.* 1st ed.: O&aapos; Reilly Community Press; 2002.
116. Python Software Foundation. *Python Language Reference, version 2.7.* Available at: <http://www.python.org>; 2015.
117. *MySQLdb User's Guide.* Available at: <http://mysql-python.sourceforge.net/MySQLdb.html>; 2015.
118. *Microsoft EXCEL 2016*: Microsoft Corporation; 2016.
119. Stark H. *Brannotate: Auswertung der optimalen Annotation von EnzymeDetector.*: Institut für Bioinformatik und Biochemie. Technische Universität Braunschweig; 2015.
120. Bentley S: **Sequencing the species pan-genome.** *Nat Rev Micro* 2009;**7**(4):258–9, doi:10.1038/nrmicro2123.
121. Waskov D. *Statistical analysis of functional gene annotation sources used by EnzymeDetector (BRENDA) and operon prediction utilizing the genetic context.* Masterthesis; Institut für Bioinformatik und Biochemie. Technische Universität Braunschweig; 2016.

122. Chaban B, Ng SY, Jarrell KF: **Archaeal habitats — from the extreme to the ordinary**. *Can. J. Microbiol.* 2006;**52**(2):73–116, doi:10.1139/w05-147.
123. Seckbach J. *Algae and Cyanobacteria in Extreme Environments*: Springer Netherlands; 2007.
124. Konings WN, Albers S, Koning S, Driessen AJM: **The cell membrane plays a crucial role in survival of bacteria and archaea in extreme environments**. *Antonie van Leeuwenhoek* 2002;**81**(1-4):61–72.
125. Thoden JB, Holden HM, Paritala H, Firestone SM: **Structural and Functional Studies of Aspergillus clavatusN5 -Carboxyaminoimidazole Ribonucleotide Synthetase**. *Biochemistry* 2010;**49**(4):752–60, doi:10.1021/bi901599u.
126. Sayed A, Ghazy MA, Ferreira AJS, Setubal JC, Chambergo FS, Ouf A, Adel M, Dawe AS, Archer JAC, Bajic VB et al.: **A novel mercuric reductase from the unique deep brine environment of Atlantis II in the Red Sea**. *The Journal of biological chemistry* 2014;**289**(3):1675–87, doi:10.1074/jbc.M113.493429.
127. ENTNER N, DOUDOROFF M: **Glucose and gluconic acid oxidation of Pseudomonas saccharophila**. *The Journal of biological chemistry* 1952;**196**(2):853–62.
128. Reher M, Fuhrer T, Bott M, Schönheit P: **The nonphosphorylative Entner-Doudoroff pathway in the thermoacidophilic euryarchaeon Picrophilus torridus involves a novel 2-keto-3-deoxygluconate- specific aldolase**. *Journal of bacteriology* 2010;**192**(4):964–74, doi:10.1128/JB.01281-09.
129. Ahmed H, Ettema TJG, Tjaden B, Geerling ACM, van der Oost J, Siebers B: **The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation**. *The Biochemical journal* 2005;**390**(Pt 2):529–40, doi:10.1042/BJ20041711.
130. van de Vossenberg, Jack L. C. M., Driessen AJM, Konings WN: **The essence of being extremophilic. The role of the unique archaeal membrane lipids**. *Extremophiles* 1998;**2**(3):163–70, doi:10.1007/s007920050056.
131. Valentine DL: **Adaptations to energy stress dictate the ecology and evolution of the Archaea**. *Nat Rev Micro* 2007;**5**(4):316–23, doi:10.1038/nrmicro1619.
132. Sharma A, Kawarabayasi Y, Satyanarayana T: **Acidophilic bacteria and archaea. Acid stable biocatalysts and their potential applications**. *Extremophiles* 2012;**16**(1):1–19, doi:10.1007/s00792-011-0402-3.
133. Harborne JB (Ed). *Phytochemical dictionary: A handbook of bioactive compounds from plants*. 2nd ed.: Taylor & Francis; 1999.
134. Littlechild JA: **Archaeal Enzymes and Applications in Industrial Biocatalysts**. *Archaea* 2015;**2015**(24):1–10, doi:10.1155/2015/147671.
135. Blum P. *Archaea: New Models for Prokaryotic Biology*. Caister Academic Press; 2008.
136. Ng SYM, Zolghadr B, Driessen AJM, Albers S, Jarrell KF: **Cell Surface Structures of Archaea**. *Journal of bacteriology* 2008;**190**(18):6039–47, doi:10.1128/JB.00546-08.
137. Offermanns S (Ed). *Encyclopedia of molecular pharmacology*. 2nd ed.: Springer; 2008.

138. Estelmann S, Hügler M, Eisenreich W, Werner K, Berg IA, Ramos-Vera WH, Say RF, Kockelkorn D, Gad'on N, Fuchs G: **Labeling and Enzyme Studies of the Central Carbon Metabolism in Metallosphaera sedula.** *Journal of bacteriology* 2011;**193**(5):1191–200, doi:10.1128/JB.01155-10.
139. Hügler M, Huber H, Stetter KO, Fuchs G: **Autotrophic CO₂ fixation pathways in archaea (Crenarchaeota).** *Archives of microbiology* 2003;**179**(3):160–73, doi:10.1007/s00203-002-0512-5.
140. Berg IA, Kockelkorn D, Buckel W, Fuchs G: **A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea.** *Science* 2007;**318**(5857):1782–6, doi:10.1126/science.1149976.
141. Berg IA, Ramos-Vera WH, Petri A, Huber H, Fuchs G: **Study of the distribution of autotrophic CO₂ fixation cycles in Crenarchaeota.** *Microbiology* 2009;**156**(1):256–69, doi:10.1099/mic.0.034298-0.
142. Bar-Even A, Noor E, Milo R: **A survey of carbon fixation pathways through a quantitative lens.** *Journal of experimental botany* 2012;**63**(6):2325–42, doi:10.1093/jxb/err417.
143. Huber G, Spinnler C, Gambacorta A, Stetter KO: **Metallosphaera sedula gen. and sp. nov. Represents a New Genus of Aerobic, Metal-Mobilizing, Thermoacidophilic Archaeobacteria.** *Systematic and Applied Microbiology* 1989;**12**(1):38–47, doi:10.1016/S0723-2020(89)80038-4.
144. Huber H, Prangishvili D: **Sulfolobales.** In: *The Prokaryotes. The Prokaryotes. Edited by Dworkin, M, Falkow, S, Rosenberg, E, Schleifer, K, Stackebrandt, E: Springer New York; 2006:23–51.*
145. Carrillo N, Ceccarelli EA: **Open questions in ferredoxin-NADP⁺ reductase catalytic mechanism.** *European journal of biochemistry / FEBS* 2003;**270**(9):1900–15.
146. Welp G, Blume HP, Scheffer F, Brümmer GW, Thiele-Bruhn S, Schachtschabel P, Horn R, Kandeler E, Kögel-Knabner I, Kretzschmar R. *Scheffer/Schachtschabel: Lehrbuch der Bodenkunde: Spektrum Akademischer Verlag; 2009.*
147. *Eintrag zu Quecksilbersulfid in der GESTIS-Stoffdatenbank des IFA.* Available at: [http://gestis.itrust.de/nxt/gateway.dll/gestis_de/004600.xml?f=templates\\$fn=default.htm\\$3.0](http://gestis.itrust.de/nxt/gateway.dll/gestis_de/004600.xml?f=templates$fn=default.htm$3.0) [accessed 17.03.2016].
148. Barkay T, Miller SM, Summers AO: **Bacterial mercury resistance from atoms to ecosystems.** *FEMS microbiology reviews* 2003;**27**(2-3):355–84.
149. Fox B, Walsh CT: **Mercuric reductase. Purification and characterization of a transposon-encoded flavoprotein containing an oxidation-reduction-active disulfide.** *The Journal of biological chemistry* 1982;**257**(5):2498–503.
150. Simbahan J, Kurth E, Schelert J, Dillman A, Moriyama E, Jovanovich S, Blum P: **Community Analysis of a Mercury Hot Spring Supports Occurrence of Domain-Specific Forms of Mercuric Reductase.** *Applied and Environmental Microbiology* 2005;**71**(12):8836–45, doi:10.1128/AEM.71.12.8836-8845.2005.
151. Schelert J, Dixit V, Hoang V, Simbahan J, Drozda M, Blum P: **Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon Sulfolobus solfataricus by use of gene disruption.** *Journal of bacteriology* 2004;**186**(2):427–37.

152. Cárdenas JP, Valdés J, Quatrini R, Duarte F, Holmes DS: **Lessons from the genomes of extremely acidophilic bacteria and archaea with special emphasis on bioleaching microorganisms.** *Appl Microbiol Biotechnol* 2010;**88**(3):605–20, doi:10.1007/s00253-010-2795-9.
153. Hou Y: **CCA addition to tRNA. Implications for tRNA quality control.** *IUBMB Life* 2010:NA-NA, doi:10.1002/iub.301.
154. Vörtler S, Mörl M: **tRNA-nucleotidyltransferases. Highly unusual RNA polymerases with vital functions.** *FEBS Letters* 2010;**584**(2):297–302, doi:10.1016/j.febslet.2009.10.078.
155. Yue D, Maizels N, Weiner AM: **CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile *Sulfolobus shibatae*.** *RNA (New York, N.Y.)* 1996;**2**(9):895–908.
156. Nomura Y, Ohno S, Nishikawa K, Yokogawa T: **Correlation between the stability of tRNA tertiary structure and the catalytic efficiency of a tRNA-modifying enzyme, archaeal tRNA-guanine transglycosylase.** *Genes Cells* 2016;**21**(1):41–52, doi:10.1111/gtc.12317.
157. Oliva R, Tramontano A, Cavallo L: **Mg²⁺ binding and archaeosine modification stabilize the G15 C48 Levitt base pair in tRNAs.** *RNA (New York, N.Y.)* 2007;**13**(9):1427–36, doi:10.1261/rna.574407.
158. Gregson JM, Crain PF, Edmonds CG, Gupta R, Hashizume T, Phillipson DW, McCloskey JA: **Structure of the archaeal transfer RNA nucleoside G*-15 (2-amino-4,7-dihydro- 4-oxo-7-beta-D-ribofuranosyl-1H-pyrrolo[2,3-d]pyrimidine-5-carboximide (archaeosine)).** *The Journal of biological chemistry* 1993;**268**(14):10076–86.
159. Watanabe M, Matsuo M, Tanaka S, Akimoto H, Asahi S, Nishimura S, Katze JR, Hashizume T, Crain PF, McCloskey JA *et al.*: **Biosynthesis of Archaeosine, a Novel Derivative of 7-Deazaguanosine Specific to Archaeal tRNA, Proceeds via a Pathway Involving Base Replacement on the tRNA Polynucleotide Chain.** *Journal of Biological Chemistry* 1997;**272**(32):20146–51, doi:10.1074/jbc.272.32.20146.
160. Edmonds CG, Crain PF, Gupta R, Hashizume T, Hocart CH, Kowalak JA, Pomerantz SC, Stetter KO, McCloskey JA: **Posttranscriptional modification of tRNA in thermophilic archaea (Archaeobacteria).** *Journal of bacteriology* 1991;**173**(10):3138–48.
161. Sorokin DY, Toshchakov SV, Kolganova TV, Kublanov IV: **Halo(natrono)archaea isolated from hypersaline lakes utilize cellulose and chitin as growth substrates.** *Front. Microbiol.* 2015;**6**(220):1, doi:10.3389/fmicb.2015.00942.
162. Johnsen U, Dambeck M, Zaiss H, Fuhrer T, Soppa J, Sauer U, Schönheit P: **D-Xylose Degradation Pathway in the Halophilic Archaeon *Haloferax volcanii*.** *Journal of Biological Chemistry* 2009;**284**(40):27290–303, doi:10.1074/jbc.M109.003814.
163. Tomme P, Warren RA, Gilkes NR: **Cellulose hydrolysis by bacteria and fungi.** *Advances in microbial physiology* 1995;**37**:1–81.
164. Baker BJ, Tyson GW, Webb RI, Flanagan J, Hugenholtz P, Allen EE, Banfield JF: **Lineages of Acidophilic Archaea Revealed by Community**

- Genomic Analysis. Science** 2006;**314**(5807):1933–5, doi:10.1126/science.1132690.
165. Stuer-Lauridsen B, Nygaard P: **Purine salvage in two halophilic archaea: characterization of salvage pathways and isolation of mutants resistant to purine analogs.** *Journal of bacteriology* 1998;**180**(3):457–63.
 166. Worrell VE, Nagle DP: **Genetic and physiological characterization of the purine salvage pathway in the archaeobacterium Methanobacterium thermoautotrophicum Marburg.** *Journal of bacteriology* 1990;**172**(6):3328–34.
 167. Sun Y, Song H, Li J, Jiang M, Li Y, Zhou J, Guo Z: **Active site binding and catalytic role of bicarbonate in 1,4-dihydroxy-2-naphthoyl coenzyme A synthases from vitamin K biosynthetic pathways.** *Biochemistry* 2012;**51**(22):4580–9, doi:10.1021/bi300486j.
 168. Shibayama K, Wachino J, Arakawa Y, Saidijam M, Rutherford NG, Henderson PJF: **Metabolism of glutamine and glutathione via gamma-glutamyltranspeptidase and glutamate transport in Helicobacter pylori: possible significance in the pathophysiology of the organism.** *Molecular microbiology* 2007;**64**(2):396–406, doi:10.1111/j.1365-2958.2007.05661.x.
 169. Feller G, Gerday C: **Psychrophilic enzymes. Hot topics in cold adaptation.** *Nat Rev Micro* 2003;**1**(3):200–8, doi:10.1038/nrmicro773.
 170. Nichols PD, Franzmann PD: **Unsaturated diether phospholipids in the Antarctic methanogen Methanococcoides burtonii.** *FEMS microbiology letters* 1992;**98**(1-3):205–8, doi:10.1111/j.1574-6968.1992.tb05515.x.
 171. Nichols DS, Miller MR, Davies NW, Goodchild A, Raftery M, Cavicchioli R: **Cold Adaptation in the Antarctic Archaeon Methanococcoides burtonii Involves Membrane Lipid Unsaturation.** *Journal of bacteriology* 2004;**186**(24):8508–15, doi:10.1128/JB.186.24.8508-8515.2004.
 172. Russell NJ, Harisson P, Johnston IA, Jaenicke R, Zuber M, Franks F, Wynn-Williams D: **Cold Adaptation of Microorganisms [and Discussion].** *Philosophical Transactions of the Royal Society B: Biological Sciences* 1990;**326**(1237):595–611, doi:10.1098/rstb.1990.0034.
 173. Chintalapati S, Kiran MD, Shivaji S: **Role of membrane lipid fatty acids in cold adaptation.** *Cellular and molecular biology (Noisy-le-Grand, France)* 2004;**50**(5):631–42.
 174. Feller G: **Protein stability and enzyme activity at extreme biological temperatures.** *J. Phys.: Condens. Matter* 2010;**22**(32):323101, doi:10.1088/0953-8984/22/32/323101.
 175. Smalås AO, Leiros HS, Os V, Willassen NP: **Cold adapted enzymes.** In: *Biotechnology annual review. Biotechnology annual review. Edited by El-Gewely, MR: Elsevier Science; 2000:1–57.*
 176. D'Amico S, Collins T, Marx J, Feller G, Gerday C: **Psychrophilic microorganisms: challenges for life.** *EMBO reports* 2006;**7**(4):385–9, doi:10.1038/sj.embor.7400662.
 177. Grogan DW: **Extreme Thermophiles.** In: *Encyclopedia of life science. Encyclopedia of life science. Edited by Cullen, KE: Facts On File; 2009.*
 178. Reed CJ, Lewis H, Trejo E, Winston V, Evilia C: **Protein Adaptations in Archaeal Extremophiles.** *Archaea* 2013;**2013**(1448):1–14, doi:10.1155/2013/373275.

179. Tomazic SJ, Klivanov AM: **Mechanisms of irreversible thermal inactivation of Bacillus alpha-amylases.** *The Journal of biological chemistry* 1988;**263**(7):3086–91.
180. Deppenmeier U: **The unique biochemistry of methanogenesis.** In: *Progress in nucleic acid research and molecular biology. Progress in nucleic acid research and molecular biology. Edited by Moldave, K: Academic Press; 2002:223–83.*
181. Anderson I, Ulrich LE, Lupa B, Susanti D, Porat I, Hooper SD, Lykidis A, Sieprawska-Lupa M, Dharmarajan L, Goltsman E *et al.*: **Genomic characterization of methanomicrobiales reveals three classes of methanogens.** *PloS one* 2009;**4**(6):e5797, doi:10.1371/journal.pone.0005797.
182. Sakai S, Conrad R, Liesack W, Imachi H: **Methanocella arvoryzae sp. nov., a hydrogenotrophic methanogen isolated from rice field soil.** *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* 2010;**60**(12):2918–23, doi:10.1099/ijs.0.020883-0.
183. Thauer RK, Kaster A, Seedorf H, Buckel W, Hedderich R: **Methanogenic archaea. Ecologically relevant differences in energy conservation.** *Nat Rev Micro* 2008;**6**(8):579–91, doi:10.1038/nrmicro1931.
184. Milkov AV: **Global estimates of hydrate-bound gas in marine sediments. How much is really out there?** *Earth-Science Reviews* 2004;**66**(3-4):183–97, doi:10.1016/j.earscirev.2003.11.002.
185. Buffett B, Archer D: **Global inventory of methane clathrate. Sensitivity to changes in the deep ocean.** *Earth and Planetary Science Letters* 2004;**227**(3-4):185–99, doi:10.1016/j.epsl.2004.09.005.
186. Daas PJ, Hagen WR, Keltjens JT, van der Drift C, Vogels GD: **Activation mechanism of methanol:5-hydroxybenzimidazolylcobamide methyltransferase from Methanosarcina barkeri.** *The Journal of biological chemistry* 1996;**271**(37):22346–51.
187. Srinivasan G: **Pyrrolysine Encoded by UAG in Archaea. Charging of a UAG-Decoding Specialized tRNA.** *Science* 2002;**296**(5572):1459–62, doi:10.1126/science.1069588.
188. Hao B: **A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase.** *Science* 2002;**296**(5572):1462–6, doi:10.1126/science.1069556.
189. Storbeck S, Rolfes S, Raux-Deery E, Warren MJ, Jahn D, Layer G: **A Novel Pathway for the Biosynthesis of Heme in Archaea. Genome-Based Bioinformatic Predictions and Experimental Evidence.** *Archaea* 2010;**2010**(13-14):1–15, doi:10.1155/2010/175050.
190. Murphy SA, Solomon BM, Meng S, Copeland JM, Shaw TJ, Ferry JL: **Geochemical Production of Reactive Oxygen Species From Biogeochemically Reduced Fe.** *Environ. Sci. Technol.* 2014;**48**(7):3815–21, doi:10.1021/es4051764.
191. Apel K, Hirt H: **Reactive oxygen species: metabolism, oxidative stress, and signal transduction.** *Annual review of plant biology* 2004;**55**:373–99, doi:10.1146/annurev.arplant.55.031903.141701.
192. Schmidt RF, Lang F. *Physiologie des Menschen: Mit Pathophysiologie.* 30th ed.: Springer Medizin Verlag Heidelberg; 2007.

193. Cannio R, Fiorentino G, Morana A, Rossi M, Bartolucci S: **Oxygen: friend or foe? Archaeal superoxide dismutases in the protection of intra- and extracellular oxidative stress.** *Frontiers in bioscience a journal and virtual library* 2000;**5**:D768-79.
194. Zhao X, Drlica K: **Reactive oxygen species and the bacterial response to lethal stress.** *Current Opinion in Microbiology* 2014;**21**:1–6, doi:10.1016/j.mib.2014.06.008.
195. Vouhier S, Mary J, Dautin N, Vinh J, Friguet B, Ladant D: **Essential role of methionine residues in calmodulin binding to Bordetella pertussis adenylate cyclase, as probed by selective oxidation and repair by the peptide methionine sulfoxide reductases.** *The Journal of biological chemistry* 2004;**279**(29):30210–8, doi:10.1074/jbc.M400604200.
196. Bolten CJ, Schröder H, Dickschat J, Wittmann C: **Towards methionine overproduction in Corynebacterium glutamicum--methanethiol and dimethyldisulfide as reduced sulfur sources.** *Journal of microbiology and biotechnology* 2010;**20**(8):1196–203.
197. Baumer S, Ide T, Jacobi C, Johann A, Gottschalk G, Deppenmeier U: **The F420H₂ dehydrogenase from Methanosarcina mazei is a Redox-driven proton pump closely related to NADH dehydrogenases.** *The Journal of biological chemistry* 2000;**275**(24):17968–73, doi:10.1074/jbc.M000650200.
198. Chastain CJ, Chollet R: **Regulation of pyruvate, orthophosphate dikinase by ADP-/Pi-dependent reversible phosphorylation in C₃ and C₄ plants.** *Plant Physiology and Biochemistry* 2003;**41**(6-7):523–32, doi:10.1016/S0981-9428(03)00065-2.
199. Burnell JN, Hatch MD: **Regulation of C₄ photosynthesis: purification and properties of the protein catalyzing ADP-mediated inactivation and Pi-mediated activation of pyruvate, Pi dikinase.** *Archives of biochemistry and biophysics* 1985;**237**(2):490–503.
200. Gonçalves LG, Borges N, Serra F, Fernandes PL, Dopazo H, Santos H: **Evolution of the biosynthesis of di-myo-inositol phosphate, a marker of adaptation to hot marine environments.** *Environmental Microbiology* 2012;**14**(3):691–701, doi:10.1111/j.1462-2920.2011.02621.x.
201. Borges N, Gonçalves LG, Rodrigues MV, Siopa F, Ventura R, Maycock C, Lamosa P, Santos H: **Biosynthetic pathways of inositol and glycerol phosphodiester used by the hyperthermophile Archaeoglobus fulgidus in stress adaptation.** *Journal of bacteriology* 2006;**188**(23):8128–35, doi:10.1128/JB.01129-06.
202. Rodrigues MV, Borges N, Henriques M, Lamosa P, Ventura R, Fernandes C, Empadinhas N, Maycock C, da Costa MS, Santos H: **Bifunctional CTP:inositol-1-phosphate cytidyltransferase/CDP-inositol:inositol-1-phosphate transferase, the key enzyme for di-myo-inositol-phosphate synthesis in several (hyper)thermophiles.** *Journal of bacteriology* 2007;**189**(15):5405–12, doi:10.1128/JB.00465-07.
203. Faria TQ, Knapp S, Ladenstein R, Maçanita AL, Santos H: **Protein stabilisation by compatible solutes: effect of mannosylglycerate on unfolding thermodynamics and activity of ribonuclease A.** *Chembiochem a European journal of chemical biology* 2003;**4**(8):734–41, doi:10.1002/cbic.200300574.

204. Lindahl T: **Instability and decay of the primary structure of DNA.** *Nature* 1993;**362**(6422):709–15, doi:10.1038/362709a0.
205. Lindahl T, Nyberg B: **Heat-induced deamination of cytosine residues in deoxyribonucleic acid.** *Biochemistry* 1974;**13**(16):3405–10, doi:10.1021/bi00713a035.
206. Tabor CW, Tabor H: **Polyamines.** *Annual review of biochemistry* 1984;**53**:749–90, doi:10.1146/annurev.bi.53.070184.003533.
207. Tabor CW, Tabor H: **Polyamines in microorganisms.** *Microbiological reviews* 1985;**49**(1):81–99.
208. Morimoto N, Fukuda W, Nakajima N, Masuda T, Terui Y, Kanai T, Oshima T, Imanaka T, Fujiwara S: **Dual Biosynthesis Pathway for Longer-Chain Polyamines in the Hyperthermophilic Archaeon Thermococcus kodakarensis.** *Journal of bacteriology* 2010;**192**(19):4991–5001, doi:10.1128/JB.00279-10.
209. Knott JM: **Biosynthesis of long-chain polyamines by crenarchaeal polyamine synthases from Hyperthermus butylicus and Pyrobaculum aerophilum.** *FEBS Letters* 2009;**583**(21):3519–24, doi:10.1016/j.febslet.2009.10.014.
210. Mine S, Ikegami T, Kawasaki K, Nakamura T, Uegaki K: **Expression, refolding, and purification of active diacetylchitobiose deacetylase from Pyrococcus horikoshii.** *Protein Expression and Purification* 2012;**84**(2):265–9, doi:10.1016/j.pep.2012.06.002.
211. Razin A, Shemer R: **DNA Methylation: Evolution.** In: *Encyclopedia of life sciences.* *Encyclopedia of life sciences*: Wiley; 2001.
212. Ehrlich M, Gama-Sosa MA, Carreira LH, Ljungdahl LG, Kuo KC, Gehrke CW: **DNA methylation in thermophilic bacteria. N 4 -methylcytosine, 5-methylcytosine, and N 5 methyladenine.** *Nucl Acids Res* 1985;**13**(4):1399–412, doi:10.1093/nar/13.4.1399.
213. Yang H, Fitz-Gibbon S, Marcotte EM, Tai JH, Hyman EC, Miller JH: **Characterization of a Thermostable DNA Glycosylase Specific for U/G and T/G Mismatches from the Hyperthermophilic Archaeon Pyrobaculum aerophilum.** *Journal of bacteriology* 2000;**182**(5):1272–9, doi:10.1128/JB.182.5.1272-1279.2000.
214. Bujnicki JM, Radlinska M: **Molecular evolution of DNA-(cytosine-N4) methyltransferases: evidence for their polyphyletic origin.** *Nucleic Acids Research* 1999;**27**(22):4501–9.
215. Squires JE, Preiss T: **Function and detection of 5-methylcytosine in eukaryotic RNA.** *Epigenomics* 2010;**2**(5):709–15, doi:10.2217/epi.10.47.
216. Lambrecht JA, Flynn JM, Downs DM: **Conserved YjgF Protein Family Deaminates Reactive Enamine/Imine Intermediates of Pyridoxal 5'-Phosphate (PLP)-dependent Enzyme Reactions.** *Journal of Biological Chemistry* 2012;**287**(5):3454–61, doi:10.1074/jbc.M111.304477.
217. Massant J, Glansdorff N: **Metabolic channelling of carbamoyl phosphate in the hyperthermophilic archaeon Pyrococcus furiosus: dynamic enzyme-enzyme interactions involved in the formation of the channelling complex.** *Biochemical Society transactions* 2004;**32**(Pt 2):306–9.

218. Ladenstein R, Ren B: **Protein disulfides and protein disulfide oxidoreductases in hyperthermophiles.** *The FEBS journal* 2006;**273**(18):4170–85, doi:10.1111/j.1742-4658.2006.05421.x.
219. Hedderich R, Hamann N, Bennati M: **Heterodisulfide reductase from methanogenic archaea: a new catalytic role for an iron-sulfur cluster.** *Biological chemistry* 2005;**386**(10):961–70, doi:10.1515/BC.2005.112.
220. Harwood CS, Gibson J: **Anaerobic and aerobic metabolism of diverse aromatic compounds by the photosynthetic bacterium *Rhodospseudomonas palustris*.** *Applied and Environmental Microbiology* 1988;**54**(3):712–7.
221. HUGH R, LEIFSON E: **The taxonomic significance of fermentative versus oxidative metabolism of carbohydrates by various gram negative bacteria.** *Journal of bacteriology* 1953;**66**(1):24–6.
222. Haferkamp P, Kutschki S, Treichel J, Hemeda H, Sewczyk K, Hoffmann D, Zaparty M, Siebers B: **An additional glucose dehydrogenase from *Sulfolobus solfataricus*. Fine-tuning of sugar degradation?** *Biochem. Soc. Trans.* 2011;**39**(1):77–81, doi:10.1042/BST0390077.
223. James KD, Bonete MJ, Byrom D, Danson MJ, Hough DW: **Citrate synthase from *Haloferax volcanii*: enzyme purification and gene cloning.** *Biochemical Society transactions* 1992;**20**(1):12S.
224. Nelson DL, Cox MM, Lehninger AL, Beginnen K, Nelson-Cox. *Lehninger Biochemie: Mit 40 Tabellen.* 3rd ed.: Springer; 2005.
225. Madigan MT, Brock TD. *Brock biology of microorganisms.* 13th ed.: Pearson; 2012.
226. Grochowski LL, Xu H, Leung K, White RH: **Characterization of an Fe(2+)-dependent archaeal-specific GTP cyclohydrolase, MptA, from *Methanocaldococcus jannaschii*.** *Biochemistry* 2007;**46**(22):6658–67, doi:10.1021/bi700052a.
227. Keltjens JT, Vogels GD: **Methanopterins and methanogenic bacteria.** *BioFactors (Oxford, England)* 1988;**1**(1):95–103.
228. White R: **Methanopterin biosynthesis. Methylation of the biosynthetic intermediates.** *Biochimica et Biophysica Acta (BBA) - General Subjects* 1998;**1380**(2):257–67, doi:10.1016/S0304-4165(97)00148-7.
229. DiMarco AA, Bobik TA, Wolfe RS: **Unusual coenzymes of methanogenesis.** *Annual review of biochemistry* 1990;**59**:355–94, doi:10.1146/annurev.bi.59.070190.002035.
230. Weiss DS, Thauer RK: **Methanogenesis and the unity of biochemistry.** *Cell* 1993;**72**(6):819–22.
231. Thauer RK: **Biochemistry of methanogenesis. A tribute to Marjory Stephenson:1998 Marjory Stephenson Prize Lecture.** *Microbiology* 1998;**144**(9):2377–406, doi:10.1099/00221287-144-9-2377.
232. Drevland RM, Jia Y, Palmer DRJ, Graham DE: **Methanogen Homaconitase Catalyzes Both Hydrolyase Reactions in Coenzyme B Biosynthesis.** *Journal of Biological Chemistry* 2008;**283**(43):28888–96, doi:10.1074/jbc.M802159200.
233. Drevland RM, Waheed A, Graham DE: **Enzymology and evolution of the pyruvate pathway to 2-oxobutyrates in *Methanocaldococcus jannaschii*.** *Journal of bacteriology* 2007;**189**(12):4391–400, doi:10.1128/JB.00166-07.

234. Greco A, Callé A, Morfin F, Thouvenot D, Cayre M, Kindbeiter K, Martin L, Levillain O, Diaz J: **S-adenosyl methionine decarboxylase activity is required for the outcome of herpes simplex virus type 1 infection and represents a new potential therapeutic target.** *FASEB journal official publication of the Federation of American Societies for Experimental Biology* 2005;**19**(9):1128–30, doi:10.1096/fj.04-2108fje.
235. Li Z, Lu S, Hou G, Ma X, Sheng D, Ni J, Shen Y: **Hjm/Hel308A DNA Helicase from Sulfolobus tokodaii Promotes Replication Fork Regression and Interacts with Hjc Endonuclease In Vitro.** *Journal of bacteriology* 2008;**190**(8):3006–17, doi:10.1128/JB.01662-07.
236. Blackwood JK, Rzechorzek NJ, Abrams AS, Maman JD, Pellegrini L, Robinson NP: **Structural and functional insights into DNA-end processing by the archaeal HerA helicase-NurA nuclease complex.** *Nucleic Acids Research* 2012;**40**(7):3183–96, doi:10.1093/nar/gkr1157.
237. Du J, Say RF, Lü W, Fuchs G, Einsle O: **Active-site remodelling in the bifunctional fructose-1,6-bisphosphate aldolase/phosphatase.** *Nature* 2011;**478**(7370):534–7, doi:10.1038/nature10458.
238. Lee YJ, Kim KJ, Kang HY, Kim H, Maeng PJ: **Involvement of GDH3 - encoded NADP + -dependent Glutamate Dehydrogenase in Yeast Cell Resistance to Stress-induced Apoptosis in Stationary Phase Cells.** *J. Biol. Chem.* 2012;**287**(53):44221–33, doi:10.1074/jbc.M112.375360.
239. Nelson MJ, Brown DP, Ferry JG: **FAD requirement for the reduction of coenzyme F420 by hydrogenase from Methanobacterium formicicum.** *Biochemical and biophysical research communications* 1984;**120**(3):775–81.
240. Lienard T, Gottschalk G: **Cloning, sequencing and expression of the genes encoding the sodium translocating N5-methyltetrahydromethanopterin. Coenzyme M methyltransferase of the methylotrophic archaeon Methanosarcina mazei Gö1.** *FEBS Letters* 1998;**425**(2):204–8.
241. Grabarse W, Mahlert F, Shima S, Thauer RK, Ermler U: **Comparison of three methyl-coenzyme M reductases from phylogenetically distant organisms: unusual amino acid modification, conservation and adaptation.** *Journal of molecular biology* 2000;**303**(2):329–44, doi:10.1006/jmbi.2000.4136.
242. Kasuya F, Hiasa M, Kawai Y, Igarashi K, Fukui M: **Inhibitory effect of quinolone antimicrobial and nonsteroidal anti-inflammatory drugs on a medium chain acyl-CoA synthetase.** *Biochemical pharmacology* 2001;**62**(3):363–7.
243. Koetsier MJ, Jekel PA, van den Berg, Marco A., Bovenberg RAL, Janssen DB: **Characterization of a phenylacetate–CoA ligase from Penicillium chrysogenum.** *Biochem. J.* 2009;**417**(2):467–76, doi:10.1042/BJ20081257.
244. Fendrihan S, Legat A, Pfaffenhuemer M, Gruber C, Weidler G, Gerbl F, Stan-Lotter H: **Extremely halophilic archaea and the issue of long-term microbial survival.** *Rev Environ Sci Biotechnol* 2006;**5**(2-3):203–18, doi:10.1007/s11157-006-0007-y.
245. Gurha P, Gupta R: **Archaeal Pus10 proteins can produce both pseudouridine 54 and 55 in tRNA.** *RNA* 2008;**14**(12):2521–7, doi:10.1261/rna.1276508.

246. Doré AS, Kilkenny ML, Jones SA, Oliver AW, Roe SM, Bell SD, Pearl LH: **Structure of an archaeal PCNA1-PCNA2-FEN1 complex: elucidating PCNA subunit and client enzyme specificity.** *Nucleic Acids Research* 2006;**34**(16):4515–26, doi:10.1093/nar/gkl623.
247. Zafrilla B, Martínez-Espinosa RM, Esclapez J, Pérez-Pomares F, Bonete MJ: **SufS protein from Haloferax volcanii involved in Fe-S cluster assembly in haloarchaea.** *Biochimica et biophysica acta* 2010;**1804**(7):1476–82, doi:10.1016/j.bbapap.2010.03.001.
248. Hidese R, Inoue T, Imanaka T, Fujiwara S: **Cysteine desulphurase plays an important role in environmental adaptation of the hyperthermophilic archaeon Thermococcus kodakarensis.** *Molecular microbiology* 2014;**93**(2):331–45, doi:10.1111/mmi.12662.
249. Beinert H: **Iron-sulfur proteins: ancient structures, still full of surprises.** *Journal of biological inorganic chemistry JBIC a publication of the Society of Biological Inorganic Chemistry* 2000;**5**(1):2–15.
250. Pickl A, Schönheit P: **The oxidative pentose phosphate pathway in the haloarchaeon Haloferax volcanii involves a novel type of glucose-6-phosphate dehydrogenase - The archaeal Zwischenferment.** *FEBS Letters* 2015;**589**(10):1105–11, doi:10.1016/j.febslet.2015.03.026.
251. Kelly MJS, Ball LJ, Krieger C, Yu Y, Fischer M, Schiffmann S, Schmieder P, Kuhne R, Bermel W, Bacher A *et al.*: **The NMR structure of the 47-kDa dimeric enzyme 3,4-dihydroxy-2-butanone-4-phosphate synthase and ligand binding studies reveal the location of the active site.** *Proceedings of the National Academy of Sciences* 2001;**98**(23):13025–30, doi:10.1073/pnas.231323598.
252. Bacher A, Eberhardt S, Fischer M, Kis K, Richter G: **BIOSYNTHESIS OF VITAMIN B₂ (RIBOFLAVIN).** *Annu. Rev. Nutr.* 2000;**20**(1):153–67, doi:10.1146/annurev.nutr.20.1.153.
253. Mashhadi Z, Zhang H, Xu H, White RH: **Identification and Characterization of an Archaeon-Specific Riboflavin Kinase.** *Journal of bacteriology* 2008;**190**(7):2615–8, doi:10.1128/JB.01900-07.
254. Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T: **Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes.** *Proceedings of the National Academy of Sciences of the United States of America* 1989;**86**(23):9355–9.
255. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA *et al.*: **Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.** *Nature* 1999;**399**(6734):323–9, doi:10.1038/20601.
256. Gribaldo S, Brochier-Armanet C: **The origin and evolution of Archaea. A state of the art.** *Philosophical Transactions of the Royal Society B: Biological Sciences* 2006;**361**(1470):1007–22, doi:10.1098/rstb.2006.1841.
257. Buchner J: **Introduction: the cellular protein folding machinery.** *Cellular and molecular life sciences CMLS* 2002;**59**(10):1587–8.
258. Large AT, Goldberg MD, Lund PA: **Chaperones and protein folding in the archaea.** *Biochem. Soc. Trans.* 2009;**37**(1):46–51, doi:10.1042/BST0370046.

259. Gutsche I, Mihalache O, Baumeister W: **ATPase cycle of an archaeal chaperonin.** *Journal of molecular biology* 2000;**300**(1):187–96, doi:10.1006/jmbi.2000.3833.
260. Yamamoto YY, Yohda M: **Thermosome: A Group II Chaperonin of Archaea.** In: *Encyclopedia of life science. Encyclopedia of life science. Edited by Cullen, KE: Facts On File; 2009:1–7.*
261. Koga Y, Morii H: **Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations.** *Microbiology and molecular biology reviews MMBR* 2007;**71**(1):97–120, doi:10.1128/MMBR.00033-06.
262. Watanabe M, Matsuo M, Tanaka S, Akimoto H, Asahi S, Nishimura S, Katze JR, Hashizume T, Crain PF, McCloskey JA *et al.*: **Biosynthesis of archaeosine, a novel derivative of 7-deazaguanosine specific to archaeal tRNA, proceeds via a pathway involving base replacement on the tRNA polynucleotide chain.** *The Journal of biological chemistry* 1997;**272**(32):20146–51.
263. Shin E, Lee J, Kim J, Jeon S, Kim Y, Nam S: **Overexpression, purification, and characterization of beta-subunit group II chaperonin from hyperthermophilic Aeropyrum pernix K1.** *Journal of microbiology and biotechnology* 2010;**20**(3):542–9.
264. Franzetti B, Schoehn G, Ebel C, Gagnon J, Ruigrok RW, Zaccai G: **Characterization of a novel complex from halophilic archaeobacteria, which displays chaperone-like activities in vitro.** *The Journal of biological chemistry* 2001;**276**(32):29906–14, doi:10.1074/jbc.M102098200.
265. Kil YV, Glazunov EA, Lanzov VA: **Characteristic thermodependence of the RadA recombinase from the hyperthermophilic archaeon Desulfurococcus amylolyticus.** *Journal of bacteriology* 2005;**187**(7):2555–7, doi:10.1128/JB.187.7.2555-2557.2005.
266. Dionne I, Brown NJ, Woodgate R, Bell SD: **On the mechanism of loading the PCNA sliding clamp by RFC.** *Mol Microbiol* 2008;**68**(1):216–22, doi:10.1111/j.1365-2958.2008.06150.x.
267. Nishida H, Tsuchiya D, Ishino Y, Morikawa K: **Overexpression, purification and crystallization of an archaeal DNA ligase from Pyrococcus furiosus.** *Acta crystallographica. Section F, Structural biology and crystallization communications* 2005;**61**(Pt 12):1100–2, doi:10.1107/S1744309105038649.
268. Mortusewicz O, Rothbauer U, Cardoso MC, Leonhardt H: **Differential recruitment of DNA Ligase I and III to DNA repair sites.** *Nucleic Acids Research* 2006;**34**(12):3523–32, doi:10.1093/nar/gkl492.
269. Kim YJ, Lee HS, Bae SS, Jeon JH, Yang SH, Lim JK, Kang SG, Kwon S, Lee J: **Cloning, expression, and characterization of a DNA ligase from a hyperthermophilic archaeon Thermococcus sp.** *Biotechnology letters* 2006;**28**(6):401–7, doi:10.1007/s10529-005-6070-6.
270. Rolland J, Gueguen Y, Persillon C, Masson J, Dietrich J: **Characterization of a thermophilic DNA ligase from the archaeon Thermococcus fumicolans.** *FEMS microbiology letters* 2004;**236**(2):267–73, doi:10.1016/j.femsle.2004.05.045.
271. Yamauchi N: **The pathway of leucine to mevalonate in halophilic archaea: efficient incorporation of leucine into isoprenoidal lipid with the involvement of isovaleryl-CoA dehydrogenase in Halobacterium**

- salinarum**. *Bioscience, biotechnology, and biochemistry* 2010;**74**(2):443–6, doi:10.1271/bbb.90814.
272. Dons L, Olsen JE, Rasmussen OF: **Characterization of two putative *Listeria monocytogenes* genes encoding polypeptides homologous to the sensor protein CheA and the response regulator CheY of chemotaxis**. *DNA sequence the journal of DNA sequencing and mapping* 1994;**4**(5):301–11.
273. Greck M, Platzer J, Sourjik V, Schmitt R: **Analysis of a chemotaxis operon in *Rhizobium meliloti***. *Molecular microbiology* 1995;**15**(6):989–1000.
274. Ward MJ, Bell AW, Hamblin PA, Packer HL, Armitage JP: **Identification of a chemotaxis operon with two cheY genes in *Rhodobacter sphaeroides***. *Molecular microbiology* 1995;**17**(2):357–66.
275. Keyzer J de, van der Sluis, Eli O, Spelbrink REJ, Nijstad N, Kruijff B de, Nouwen N, van der Does C, Driessen AJM: **Covalently dimerized SecA is functional in protein translocation**. *The Journal of biological chemistry* 2005;**280**(42):35255–60, doi:10.1074/jbc.M506157200.
276. Kim DY, Kim KK: **The Structural Basis for the Activation and Peptide Recognition of Bacterial ClpP**. *Journal of molecular biology* 2008;**379**(4):760–71, doi:10.1016/j.jmb.2008.04.036.
277. Polz MF, Alm EJ, Hanage WP: **Horizontal gene transfer and the evolution of bacterial and archaeal population structure**. *Trends in Genetics* 2013;**29**(3):170–5, doi:10.1016/j.tig.2012.12.006.
278. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA *et al.*: **Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima***. *Nature* 1999;**399**(6734):323–9, doi:10.1038/20601.
279. Ignatov KB, Kramarov VM: **DNA ligases from thermophilic bacteria enhance PCR amplification of long DNA sequences**. *Biochemistry. Biokhimiia* 2009;**74**(5):557–61.
280. Garcia-Horsman JA, Puustinen A, Gennis RB, Wikström M: **Proton transfer in cytochrome bo3 ubiquinol oxidase of *Escherichia coli*: second-site mutations in subunit I that restore proton pumping in the mutant Asp135--Asn**. *Biochemistry* 1995;**34**(13):4428–33.
281. Zeng J, Wu L, Zhang X, Liu Y, Deng G, Li D: **Oct-2-en-4-ynoyl-CoA as a specific inhibitor of acyl-CoA oxidase**. *Organic letters* 2008;**10**(19):4287–90, doi:10.1021/ol801571n.
282. Vinokur JM, Korman TP, Cao Z, Bowie JU: **Evidence of a Novel Mevalonate Pathway in Archaea**. *Biochemistry* 2014;**53**(25):4161–8, doi:10.1021/bi500566q.
283. Jordan SW, Cronan JE: **A new metabolic link. The acyl carrier protein of lipid synthesis donates lipoic acid to the pyruvate dehydrogenase complex in *Escherichia coli* and mitochondria**. *The Journal of biological chemistry* 1997;**272**(29):17903–6.
284. Gomes M, Negrato C: **Alpha-lipoic acid as a pleiotropic compound with potential therapeutic use in diabetes and other chronic diseases**. *Diabetol Metab Syndr* 2014;**6**(1):80, doi:10.1186/1758-5996-6-80.

285. Gorąca A, Huk-Kolega H, Piechota A, Kleniewska P, Ciejka E, Skibska B: **Lipoic acid - biological activity and therapeutic potential.** *Pharmacological reports PR* 2011;**63**(4):849–58.
286. Park Y, Yoon S, Lee H: **A novel diene lactone hydrolase from the thermoacidophilic archaeon *Sulfolobus solfataricus* P1: purification, characterization, and expression.** *Biochimica et biophysica acta* 2010;**1800**(11):1164–72, doi:10.1016/j.bbagen.2010.07.006.
287. Kaschabek SR, Kuhn B, Müller D, Schmidt E, Reineke W: **Degradation of aromatics and chloroaromatics by *Pseudomonas* sp. strain B13: purification and characterization of 3-oxoadipate:succinyl-coenzyme A (CoA) transferase and 3-oxoadipyl-CoA thiolase.** *Journal of bacteriology* 2002;**184**(1):207–15.
288. Schlömann M: **Evolution of chlorocatechol catabolic pathways. Conclusions to be drawn from comparisons of lactone hydrolases.** *Biodegradation* 1994;**5**(3-4):301–21.
289. Acharya P, Goenrich M, Hagemeyer CH, Demmer U, Vorholt JA, Thauer RK, Ermler U: **How an enzyme binds the C1 carrier tetrahydromethanopterin. Structure of the tetrahydromethanopterin-dependent formaldehyde-activating enzyme (Fae) from *Methylobacterium extorquens* AM1.** *The Journal of biological chemistry* 2005;**280**(14):13712–9, doi:10.1074/jbc.M412320200.
290. Lobo AL, Zinder SH: **Nitrogenase in the archaeobacterium *Methanosarcina barkeri* 227.** *Journal of bacteriology* 1990;**172**(12):6789–96.
291. Hassan YI, Zemleni J: **Epigenetic regulation of chromatin structure and gene function by biotin.** *The Journal of nutrition* 2006;**136**(7):1763–5.
292. Kamachi T, Kouno T, Doitomi K, Yoshizawa K: **Generation of adenosyl radical from S-adenosylmethionine (SAM) in biotin synthase.** *Journal of Inorganic Biochemistry* 2011;**105**(6):850–7, doi:10.1016/j.jinorgbio.2011.03.013.
293. Tipton PA, Peisach J: **Characterization of the multiple catalytic activities of tartrate dehydrogenase.** *Biochemistry* 1990;**29**(7):1749–56.
294. Furuyoshi S, Nawa Y, Kawabata N, Tanaka H, Soda K: **Purification and characterization of a new NAD(+)-dependent enzyme, L-tartrate decarboxylase, from *Pseudomonas* sp. group Ve-2.** *Journal of biochemistry* 1991;**110**(4):520–5.
295. Giffhorn F, Kuhn A: **Purification and characterization of a bifunctional L-(+)-tartrate dehydrogenase-D-(+)-malate dehydrogenase (decarboxylating) from *Rhodopseudomonas sphaeroides* Y.** *Journal of bacteriology* 1983;**155**(1):281–90.
296. Jeon WY, Yoon BH, Ko BS, Shim WY, Kim JH: **Xylitol production is increased by expression of codon-optimized *Neurospora crassa* xylose reductase gene in *Candida tropicalis*.** *Bioprocess Biosyst Eng* 2012;**35**(1-2):191–8, doi:10.1007/s00449-011-0618-8.
297. Farmer JJ, Eagon RG: **Aldohexuronic acid catabolism by a soil *Aeromonas*.** *Journal of bacteriology* 1969;**97**(1):97–106.
298. Pelz A, Wieland K, Putzbach K, Hentschel P, Albert K, Götz F: **Structure and biosynthesis of staphyloxanthin from *Staphylococcus aureus*.** *The*

- Journal of biological chemistry* 2005;**280**(37):32493–8, doi:10.1074/jbc.M505070200.
299. Gotz, Windhovel, Boger, Sandmann: **Protection of photosynthesis against ultraviolet-B radiation by carotenoids in transformants of the cyanobacterium synechococcus PCC7942.** *Plant physiology* 1999;**120**(2):599–604.
300. Stahl W, Sies H: **Carotenoids and Protection against Solar UV Radiation.** *Skin Pharmacol Appl Skin Physiol* 2002;**15**(5):291–6, doi:10.1159/000064532.
301. Duursma EK, Dawson R. *Marine organic chemistry: Evolution, composition, interactions and chemistry of organic matter in seawater.* Elsevier; 1981.
302. Huang J, Xin Y, Zhang W: **Isolation, characterization and identification of a Paracoccus sp. 2-haloacid-degrading bacterium from the marine sponge Hymeniacidon perlevis.** *Journal of basic microbiology* 2011;**51**(3):318–24, doi:10.1002/jobm.201000205.
303. Okamoto S, Eltis LD: **Purification and characterization of a novel nitrile hydratase from Rhodococcus sp. RHA1.** *Molecular microbiology* 2007;**65**(3):828–38, doi:10.1111/j.1365-2958.2007.05834.x.
304. Sergiev PV, Lesnyak DV, Bogdanov AA, Dontsova OA: **Identification of Escherichia coli m2G methyltransferases: II. The yggO gene encodes a methyltransferase specific for G1835 of the 23 S rRNA.** *Journal of molecular biology* 2006;**364**(1):26–31, doi:10.1016/j.jmb.2006.09.008.
305. Osterman IA, Sergiev PV, Tsvetkov PO, Makarov AA, Bogdanov AA, Dontsova OA: **Methylated 23S rRNA nucleotide m2G1835 of Escherichia coli ribosome facilitates subunit association.** *Biochimie* 2011;**93**(4):725–9, doi:10.1016/j.biochi.2010.12.016.
306. Reshetnikov AS, Khmelenina VN, Mustakhimov II, Kalyuzhnaya M, Lidstrom M, Trotsenko YA: **Diversity and phylogeny of the ectoine biosynthesis genes in aerobic, moderately halophilic methylotrophic bacteria.** *Extremophiles life under extreme conditions* 2011;**15**(6):653–63, doi:10.1007/s00792-011-0396-x.
307. Ashida H, Saito Y, Nakano T, Tandau de Marsac N, Sekowska A, Danchin A, Yokota A: **RuBisCO-like proteins as the enolase enzyme in the methionine salvage pathway: functional and evolutionary relationships between RuBisCO-like proteins and photosynthetic RuBisCO.** *Journal of experimental botany* 2008;**59**(7):1543–54, doi:10.1093/jxb/ern104.
308. Gulmezian M, Hyman KR, Marbois BN, Clarke CF, Javor GT: **The role of UbiX in Escherichia coli coenzyme Q biosynthesis.** *Archives of biochemistry and biophysics* 2007;**467**(2):144–53, doi:10.1016/j.abb.2007.08.009.
309. Lee J, Sperandio V, Frantz DE, Longgood J, Camilli A, Phillips MA, Michael AJ: **An alternative polyamine biosynthetic pathway is widespread in bacteria and essential for biofilm formation in Vibrio cholerae.** *The Journal of biological chemistry* 2009;**284**(15):9899–907, doi:10.1074/jbc.M900110200.
310. Krzycki JA: **Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases.** *Current opinion in chemical biology* 2004;**8**(5):484–91, doi:10.1016/j.cbpa.2004.08.012.

311. Damm E, Helmke E, Thoms S, Schauer U, Nöthig E, Bakker K, Kiene RP: **Methane production in aerobic oligotrophic surface water in the central Arctic Ocean.** *Biogeosciences* 2010;**7**(3):1099–108, doi:10.5194/bg-7-1099-2010.
312. Qian Y, Tabita FR: **A global signal transduction system regulates aerobic and anaerobic CO₂ fixation in *Rhodobacter sphaeroides*.** *Journal of bacteriology* 1996;**178**(1):12–8.
313. Buenger J, Driller H: **Ectoin. An Effective Natural Substance to Prevent UVA-Induced Premature Photoaging.** *Skin Pharmacol Physiol* 2004;**17**(5):232–7, doi:10.1159/000080216.
314. Müller V: **Bacterial Fermentation.** In: *Encyclopedia of life sciences. Encyclopedia of life sciences*: Wiley; 2001.
315. Catalanotti C, Yang W, Posewitz MC, Grossman AR: **Fermentation metabolism and its evolution in algae.** *Front. Plant Sci.* 2013;**4**, doi:10.3389/fpls.2013.00150.
316. LU Z, HEGEMANN W: **ANAEROBIC TOXICITY AND BIODEGRADATION OF FORMALDEHYDE IN BATCH CULTURES.** *Water Research* 1998;**32**(1):209–15, doi:10.1016/S0043-1354(97)00181-4.
317. Omil F, Méndez D, Vidal G, Méndez R, Lema J: **Biodegradation of formaldehyde under anaerobic conditions.** *Enzyme and Microbial Technology* 1999;**24**(5-6):255–62, doi:10.1016/S0141-0229(98)00119-7.
318. Berg JM, Tymoczko JL, Stryer L, Häcker B. *Biochemie*. 6th ed.: Spektrum Akad. Verl; 2010.
319. Hamerly T, Tripet B, Wurch L, Hettich RL, Podar M, Bothner B, Copié V: **Characterization of Fatty Acids in Crenarchaeota by GC-MS and NMR.** *Archaea* 2015;**2015**(9):1–9, doi:10.1155/2015/472726.
320. Dibrova DV, Galperin MY, Mulkidjanian AY: **Phylogenomic reconstruction of archaeal fatty acid metabolism.** *Environ Microbiol* 2014;**16**(4):907–18, doi:10.1111/1462-2920.12359.
321. Nandi DL, Horowitz PM, Westley J: **Rhodanese as a thioredoxin oxidase.** *The international journal of biochemistry & cell biology* 2000;**32**(4):465–73.
322. Reimann J, Esser D, Orell A, Amman F, Pham TK, Noirel J, Lindås A, Bernander R, Wright PC, Siebers B *et al.*: **Archaeal Signal Transduction. Impact of Protein Phosphatase Deletions on Cell Size, Motility, and Energy Metabolism in *Sulfolobus acidocaldarius*.** *Mol Cell Proteomics* 2013;**12**(12):3908–23, doi:10.1074/mcp.M113.027375.
323. Matsumi R, Atomi H, Driessen AJ, van der Oost J: **Isoprenoid biosynthesis in Archaea – Biochemical and evolutionary implications.** *Research in Microbiology* 2011;**162**(1):39–52, doi:10.1016/j.resmic.2010.10.003.
324. Parlar H, Angerhöfer D. *Chemische Ökotoxikologie*. 2nd ed.: Springer; 1995.
325. Munnecke DM, Hsieh DP: **Microbial decontamination of parathion and p-nitrophenol in aqueous media.** *Applied microbiology* 1974;**28**(2):212–7.
326. Pakala SB, Gorla P, Pinjari AB, Krovdi RK, Baru R, Yanamandra M, Merrick M, Siddavattam D: **Biodegradation of methyl parathion and p-nitrophenol: evidence for the presence of a p-nitrophenol 2-hydroxylase**

- in a Gram-negative *Serratia* sp. strain DS001. *Applied microbiology and biotechnology* 2007;**73**(6):1452–62, doi:10.1007/s00253-006-0595-z.
327. Elias M, Tawfik DS: **Divergence and Convergence in Enzyme Evolution. Parallel Evolution of Paraoxonases from Quorum-quenching Lactonases.** *Journal of Biological Chemistry* 2011;**287**(1):11–20, doi:10.1074/jbc.R111.257329.
328. Du Q, Wang H, Xie J: **Thiamin (vitamin B1) biosynthesis and regulation: a rich source of antimicrobial drug targets?** *International journal of biological sciences* 2011;**7**(1):41–52.
329. Begley TP: **The biosynthesis and degradation of thiamin (vitamin B1).** *Nat. Prod. Rep.* 1996;**13**(3):177, doi:10.1039/np9961300177.
330. Hwang S, Cordova B, Chavarria N, Elbanna D, McHugh S, Rojas J, Pfeiffer F, Maupin-Furlow JA: **Conserved active site cysteine residue of archaeal THI4 homolog is essential for thiamine biosynthesis in *Haloferax volcanii*.** *BMC Microbiol* 2014;**14**(1):335, doi:10.1186/s12866-014-0260-0.
331. Vetter J: **Plant cyanogenic glycosides.** *Toxicon* 2000;**38**(1):11–36, doi:10.1016/S0041-0101(99)00128-2.
332. Wang H, Squina F, Segato F, Mort A, Lee D, Pappan K, Prade R: **High-Temperature Enzymatic Breakdown of Cellulose.** *Applied and Environmental Microbiology* 2011;**77**(15):5199–206, doi:10.1128/AEM.00199-11.
333. Maruyama A, Saito K, Ishizawa K: **Beta-cyanoalanine synthase and cysteine synthase from potato: molecular cloning, biochemical characterization, and spatial and hormonal regulation.** *Plant molecular biology* 2001;**46**(6):749–60.
334. Borup B, Ferry JG: **Cysteine biosynthesis in the Archaea: *Methanosarcina thermophila* utilizes O-acetylserine sulfhydrylase.** *FEMS microbiology letters* 2000;**189**(2):205–10.
335. Ishikawa K, Mino K, Nakamura T: **New function and application of the cysteine synthase from archaea.** *Biochemical Engineering Journal* 2010;**48**(3):315–22, doi:10.1016/j.bej.2009.10.015.
336. Deshkar A, Dhamorikar N, Godbole S, Krishnamurthi K, Saravanadevi S, Vijay R, Kaul S, Chakrabarti T: **Bioremediation of soil contaminated with organic compounds with special reference to acrylonitrile.** *Annali di chimica* 2003;**93**(9-10):729–37.
337. Kobayashi M, Suzuki T, Fujita T, Masuda M, Shimizu S: **Occurrence of enzymes involved in biosynthesis of indole-3-acetic acid from indole-3-acetonitrile in plant-associated bacteria, *Agrobacterium* and *Rhizobium*.** *Proceedings of the National Academy of Sciences of the United States of America* 1995;**92**(3):714–8.
338. Kobayashi M, Izui H, Nagasawa T, Yamada H: **Nitrilase in biosynthesis of the plant hormone indole-3-acetic acid from indole-3-acetonitrile: cloning of the *Alcaligenes* gene and site-directed mutagenesis of cysteine residues.** *Proceedings of the National Academy of Sciences of the United States of America* 1993;**90**(1):247–51.
339. Geng H, Belas R: **Molecular mechanisms underlying roseobacter-phytoplankton symbioses.** *Current Opinion in Biotechnology* 2010;**21**(3):332–8, doi:10.1016/j.copbio.2010.03.013.

340. Kawai S, Mori S, Mukai T, Murata K: **Cytosolic NADP phosphatases I and II from *Arthrobacter* sp. strain KM: implication in regulation of NAD⁺/NADP⁺ balance.** *Journal of basic microbiology* 2004;**44**(3):185–96, doi:10.1002/jobm.200310362.
341. Kawai S, Murata K: **Structure and function of NAD kinase and NADP phosphatase: key enzymes that regulate the intracellular balance of NAD(H) and NADP(H).** *Bioscience, biotechnology, and biochemistry* 2008;**72**(4):919–30, doi:10.1271/bbb.70738.
342. Spaans SK, Weusthuis RA, van der Oost J, Kengen SWM: **NADPH-generating systems in bacteria and archaea.** *Front. Microbiol.* 2015;**6**(Pt 2):159, doi:10.3389/fmicb.2015.00742.
343. Riedel T, Fiebig A, Han J, Huntemann M, Spring S, Petersen J, Ivanova NN, Markowitz V, Woyke T, Göker M *et al.*: **Genome sequence of the *Wenxinia marina* type strain (DSM 24838T), a representative of the *Roseobacter* group isolated from oilfield sediments.** *Stand. Genomic Sci.* 2014;**9**(3):855–65, doi:10.4056/sigs.5601028.
344. *Pharmakotherapie bei Haus- und Nutztieren: 99 Tabellen.* 7th ed.: Parey; 2006.
345. Breazeale SD, Ribeiro AA, McClerren AL, Raetz CRH: **A Formyltransferase Required for Polymyxin Resistance in *Escherichia coli* and the Modification of Lipid A with 4-Amino-4-deoxy-L-arabinose. IDENTIFICATION AND FUNCTION OF UDP-4-DEOXY-4-FORMAMIDO-L-ARABINOSE.** *Journal of Biological Chemistry* 2005;**280**(14):14154–67, doi:10.1074/jbc.M414265200.
346. Iturriaga G, Suárez R, Nova-Franco B: **Trehalose Metabolism. From Osmoprotection to Signaling.** *IJMS* 2009;**10**(9):3793–810, doi:10.3390/ijms10093793.
347. Tropis M: **The Crucial Role of Trehalose and Structurally Related Oligosaccharides in the Biosynthesis and Transfer of Mycolic Acids in *Corynebacterineae*.** *Journal of Biological Chemistry* 2005;**280**(28):26573–85, doi:10.1074/jbc.M502104200.
348. Argüelles JC: **Physiological roles of trehalose in bacteria and yeasts: a comparative analysis.** *Archives of microbiology* 2000;**174**(4):217–24.
349. Woodruff PJ, Carlson BL, Siridechadilok B, Pratt MR, Senaratne RH, Mougous JD, Riley LW, Williams SJ, Bertozzi CR: **Trehalose Is Required for Growth of *Mycobacterium smegmatis*.** *Journal of Biological Chemistry* 2004;**279**(28):28835–43, doi:10.1074/jbc.M313103200.
350. Bender RA: **Regulation of the Histidine Utilization (Hut) System in Bacteria.** *Microbiology and Molecular Biology Reviews* 2012;**76**(3):565–84, doi:10.1128/MMBR.00014-12.
351. Gao YS, Alvarez C, Nelson DS, Sztul E: **Molecular cloning, characterization, and dynamics of rat formiminotransferase cyclodeaminase, a Golgi-associated 58-kDa protein.** *The Journal of biological chemistry* 1998;**273**(50):33825–34.
352. Fondi M, Emiliani G, Liò P, Gribaldo S, Fani R: **The Evolution of Histidine Biosynthesis in Archaea. Insights into the his Genes Structure and Organization in LUCA.** *J Mol Evol* 2009;**69**(5):512–26, doi:10.1007/s00239-009-9286-6.

353. Boschi-Muller S, Olry A, Antoine M, Branlant G: **The enzymology and biochemistry of methionine sulfoxide reductases.** *Biochimica et biophysica acta* 2005;**1703**(2):231–8, doi:10.1016/j.bbapap.2004.09.016.
354. Moskovitz J, Singh VK, Requena J, Wilkinson BJ, Jayaswal RK, Stadtman ER: **Purification and characterization of methionine sulfoxide reductases from mouse and Staphylococcus aureus and their substrate stereospecificity.** *Biochemical and biophysical research communications* 2002;**290**(1):62–5, doi:10.1006/bbrc.2001.6171.
355. Clarke DJ, Ortega XP, Mackay CL, Valvano MA, Govan JRW, Campopiano DJ, Langridge-Smith P, Brown AR: **Subdivision of the Bacterioferritin Comigratory Protein Family of Bacterial Peroxiredoxins Based on Catalytic Activity.** *Biochemistry* 2010;**49**(6):1319–30, doi:10.1021/bi901703m.
356. Nolden L, Ngouoto-Nkili CE, Bendt AK, Krämer R, Burkovski A: **Sensing nitrogen limitation in Corynebacterium glutamicum: the role of glnK and glnD.** *Molecular microbiology* 2001;**42**(5):1281–95.
357. Roberts DL, Bennett DW, Forst SA: **Identification of the site of phosphorylation on the osmosensor, EnvZ, of Escherichia coli.** *The Journal of biological chemistry* 1994;**269**(12):8728–33.
358. Pflock M, Dietz P, Schär J, Beier D: **Genetic evidence for histidine kinase HP165 being an acid sensor of Helicobacter pylori.** *FEMS microbiology letters* 2004;**234**(1):51–61, doi:10.1016/j.femsle.2004.03.023.
359. Xue J, Wang A, Zhou C, Shen S: **Penicillin acylase immobilization depending on macromolecular crowding and catalysis in aqueous-organic medium.** *Bioprocess and biosystems engineering* 2009;**32**(6):765–72, doi:10.1007/s00449-009-0301-5.
360. Sirko A, Zatyka M, Sadowy E, Hulanicka D: **Sulfate and thiosulfate transport in Escherichia coli K-12: evidence for a functional overlapping of sulfate- and thiosulfate-binding proteins.** *Journal of bacteriology* 1995;**177**(14):4134–6.
361. Pahwa S, Kaur S, Jain R, Roy N: **Structure based design of novel inhibitors for histidinol dehydrogenase from Geotrichum candidum.** *Bioorganic & medicinal chemistry letters* 2010;**20**(13):3972–6, doi:10.1016/j.bmcl.2010.04.116.
362. Hemmi H, Yamashita S, Shimoyama T, Nakayama T, Nishino T: **Cloning, expression, and characterization of cis-polyprenyl diphosphate synthase from the thermoacidophilic archaeon Sulfolobus acidocaldarius.** *Journal of bacteriology* 2001;**183**(1):401–4, doi:10.1128/JB.183.1.401-404.2001.
363. Hatori Y, Majima E, Tsuda T, Toyoshima C: **Domain organization and movements in heavy metal ion pumps: papain digestion of CopA, a Cu+-transporting ATPase.** *The Journal of biological chemistry* 2007;**282**(35):25213–21, doi:10.1074/jbc.M703520200.
364. Völlmecke C, Kötting C, Gerwert K, Lübbers M: **Spectroscopic investigation of the reaction mechanism of CopB-B, the catalytic fragment from an archaeal thermophilic ATP-driven heavy metal transporter.** *The FEBS journal* 2009;**276**(21):6172–86, doi:10.1111/j.1742-4658.2009.07320.x.

365. Angerer A, Klupp B, Braun V: **Iron transport systems of *Serratia marcescens*.** *Journal of bacteriology* 1992;**174**(4):1378–87.
366. Leopoldseder S, Hettwer S, Sterner R: **Evolution of multi-enzyme complexes: the case of tryptophan synthase.** *Biochemistry* 2006;**45**(47):14111–9, doi:10.1021/bi061684b.
367. Yasutake Y, Yao M, Sakai N, Kirita T, Tanaka I: **Crystal structure of the *Pyrococcus horikoshii* isopropylmalate isomerase small subunit provides insight into the dual substrate specificity of the enzyme.** *Journal of molecular biology* 2004;**344**(2):325–33, doi:10.1016/j.jmb.2004.09.035.
368. Sugiyama T, Uchida C, Oda T, Kitagawa M, Hayashi H, Ichiyama A: **Involvement of CCAAT/enhancer-binding protein in regulation of the rat serine:pyruvate/alanine:glyoxylate aminotransferase gene expression.** *FEBS Letters* 2001;**508**(1):16–22.
369. Waechter CJ, Lennarz WJ: **The Role of Polyprenol-Linked Sugars in Glycoprotein Synthesis.** *Annu. Rev. Biochem.* 1976;**45**(1):95–112, doi:10.1146/annurev.bi.45.070176.000523.
370. Calo D, Guan Z, Naparstek S, Eichler J: **Different routes to the same ending. Comparing the N-glycosylation processes of *Haloferax volcanii* and *Haloarcula marismortui*, two halophilic archaea from the Dead Sea.** *Molecular microbiology* 2011;**81**(5):1166–77, doi:10.1111/j.1365-2958.2011.07781.x.
371. Calo D, Kaminski L, Eichler J: **Protein glycosylation in Archaea. Sweet and extreme.** *Glycobiology* 2010;**20**(9):1065–76, doi:10.1093/glycob/cwq055.
372. Cohen-Rosenzweig C, Yurist-Doutsch S, Eichler J: **AgIS, a Novel Component of the *Haloferax volcanii* N-Glycosylation Pathway, Is a Dolichol Phosphate-Mannose Mannosyltransferase.** *Journal of bacteriology* 2012;**194**(24):6909–16, doi:10.1128/JB.01716-12.
373. Calo D, Guan Z, Eichler J: **Glyco-engineering in Archaea. Differential N-glycosylation of the S-layer glycoprotein in a transformed *Haloferax volcanii* strain.** *Microbial Biotechnology* 2011;**4**(4):461–70, doi:10.1111/j.1751-7915.2011.00250.x.
374. Fazekas D, Koltai M, Türei D, Módos D, Pálffy M, Dúl Z, Zsákai L, Szalay-Bekő M, Lenti K, Farkas IJ *et al.*: **Signalink 2 – a signaling pathway resource with multi-layered regulatory networks.** *BMC Syst Biol* 2013;**7**(1):7, doi:10.1186/1752-0509-7-7.
375. Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM *et al.*: **MODOMICS. A database of RNA modification pathways--2013 update.** *Nucleic Acids Research* 2012;**41**(D1):D262-D267, doi:10.1093/nar/gks1007.
376. Rawlings ND, Waller M, Barrett AJ, Bateman A: **MEROPS. The database of proteolytic enzymes, their substrates and inhibitors.** *Nucl. Acids Res.* 2013;**42**(D1):D503-D509, doi:10.1093/nar/gkt953.
377. F. DeLong, Norman R. Pace, Edward: **Environmental Diversity of Bacteria and Archaea.** *Systematic Biology* 2001;**50**(4):470–8, doi:10.1080/10635150118513.
378. Croft MT, Warren MJ, Smith AG: **Algae Need Their Vitamins.** *Eukaryotic Cell* 2006;**5**(8):1175–83, doi:10.1128/EC.00097-06.

379. Giovannoni SJ: **Vitamins in the sea.** *Proceedings of the National Academy of Sciences* 2012;**109**(35):13888–9, doi:10.1073/pnas.1211722109.
380. Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F. *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*: Imprint Springer; 2014.
381. Munoz R, Yarza P, Ludwig W, Euzéby J, Amann R, Schleifer K, Oliver Glöckner F, Rosselló-Móra R: **Release LTPs104 of the All-Species Living Tree.** *Systematic and Applied Microbiology* 2011;**34**(3):169–70, doi:10.1016/j.syapm.2011.03.001.
382. Newton RJ, Griffin LE, Bowles KM, Meile C, Gifford S, Givens CE, Howard EC, King E, Oakley CA, Reisch CR *et al.*: **Genome characteristics of a generalist marine bacterial lineage.** *ISME J* 2010;**4**(6):784–98, doi:10.1038/ismej.2009.150.
383. Gifford SM, Sharma S, Moran MA: **Linking activity and function to ecosystem dynamics in a coastal bacterioplankton community.** *Front. Microbiol.* 2014;**5**:1403, doi:10.3389/fmicb.2014.00185.
384. Gosink JJ, Herwig RP, Staley JT: **Octadecabacter arcticus gen. nov., sp. nov., and O. antarcticus, sp. nov., Nonpigmented, Psychrophilic Gas Vacuolate Bacteria from Polar Sea Ice and Water.** *Systematic and Applied Microbiology* 1997;**20**(3):356–65, doi:10.1016/S0723-2020(97)80003-3.
385. Lucena T, Ruvira MA, Garay E, Macian MC, Arahal DR, Pujalte MJ: **Actibacterium mucosum gen. nov., sp. nov., a marine alphaproteobacterium from Mediterranean seawater.** *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* 2012;**62**(Pt 12):2858–64, doi:10.1099/ijs.0.038026-0.
386. Fiebig A, Pradella S, Petersen J, Päuker O, Michael V, Lünsdorf H, Göker M, Klenk H, Wagner-Döbler I: **Genome of the R-body producing marine alphaproteobacterium Labrenzia alexandrii type strain (DFL-11T).** *Stand. Genomic Sci.* 2013;**7**(3):413–26, doi:10.4056/sigs.3456959.
387. Aravind L, Tatusov RL, Wolf YI, Walker D, Koonin EV: **Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles.** *Trends in Genetics* 1998;**14**(11):442–4, doi:10.1016/S0168-9525(98)01553-4.
388. Nesbo CL, L'Haridon S, Stetter KO, Doolittle WF: **Phylogenetic analyses of two "archaeal" genes in thermotoga maritima reveal multiple transfers between archaea and bacteria.** *Molecular biology and evolution* 2001;**18**(3):362–75.
389. Li CL, Jiang YT, Liu DL, Qian J, Liang JP, Shu R: **Prevalence and quantification of the uncommon Archaea phylotype Thermoplasmata in chronic periodontitis.** *Archives of Oral Biology* 2014;**59**(8):822–8, doi:10.1016/j.archoralbio.2014.05.011.
390. Woese CR, Kandler O, Wheelis ML: **Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.** *Proceedings of the National Academy of Sciences of the United States of America* 1990;**87**(12):4576–9.
391. Gunde-Cimerman N, Oren A, Plemenitaš A, Plemenitaš A. *Adaptation to Life at High Salt Concentrations in Archaea, Bacteria, and Eukarya*: Springer Netherlands; 2005.

- 392. Sato T, Atomi H, Imanaka T: **Archaeal type III RuBisCOs function in a pathway for AMP metabolism.** *Science (New York, N.Y.)* 2007;**315**(5814):1003–6, doi:10.1126/science.1135999.
- 393. Vorholt, Hafenbradl, Stetter, Thauer: **Pathways of autotrophic CO₂ fixation and of dissimilatory nitrate reduction to N₂O in *Ferroglobus placidus*.** *Archives of microbiology* 1997;**167**(1):19–23.
- 394. Anderson I, Risso C, Holmes D, Lucas S, Copeland A, Lapidus A, Cheng J, Bruce D, Goodwin L, Pitluck S *et al.*: **Complete genome sequence of *Ferroglobus placidus* AEDII12DO.** *Stand. Genomic Sci.* 2011;**5**(1):50–60, doi:10.4056/sigs.2225018.

Lebenslauf

Diplom-Biologe univ. Marcus Ulbrich

Persönliche Daten

Name	Marcus Ulbrich
Adresse	Magdeburgstraße 30A D-38124 Braunschweig
Telefon	0176 620 10 549
eMail	m.ulbrich@tu-bs.de
Nationalität	deutsch
Geburtsdatum	25.06.1986

Beruflicher Werdegang

ab 08.2012	Technische Universität Carolo-Wilhelmina zu Braunschweig, Wissenschaftlicher Mitarbeiter/Doktorand: Prof. Dr. Schomburg, Bioinformatik und Biochemie
01.2012 - 03.2012	Julius-Maximilians-Universität Würzburg, HiWi-Angestellter: Prof. Dr. Schultz, Gen-Annotations-Projekt: Tardigrada
11.2011 - 01.2012	Julius-Maximilians-Universität Würzburg, HiWi-Angestellter: Prof. Dr. Schultz, Publikation: FamilyFinder
2005 - 2006	Zivildienst Tierheim Arnstadt

Studium/Schulbildung

2010 - 2011	Julius-Maximilians-Universität Würzburg, Diplomand Lehrstuhl der Bioinformatik, Abschluss: Diplom-Biologe univ.
2006 - 2010	Julius-Maximilians-Universität Würzburg, Studium Diplom Biologie: HF Bioinformatik NF Tierökologie/Tropenbiologie, Biotechnologie
2001 - 2005	Herdergymnasium, Arnstadt Abschluss: Abitur
1997 - 2001	Herdergymnasium, Stadtilm
1993 - 1997	Grundschule, Stadtilm